SERIES A-2

# HEALTH STATISTICS

*FROM THE U.S. NATIONAL HEALTH SURVEY*



## the statistical design of the HEALTH HOUSEHOLD-INTERVIEW SURVEY


U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE


The Library of Congress catalog card for this publication is shown on the inside back cover.

# HEALTH STATISTICS

*FROM THE U.S. NATIONAL HEALTH SURVEY*

## the statistical design of the Health Household-Interview Survey

by staff of the
U. S. National Health Survey
and the
Bureau of the Census

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Marion B. Folsom, Secretary

Public Health Service
Leroy E. Burney, Surgeon General

Division of Public Health Methods
William H. Stewart, M. D., Chief

Washington, D. C. July 1958

# U. S. NATIONAL HEALTH SURVEY

Forrest E. Linder, Ph. D., Director
Theodore D. Woolsey, Assistant Director
Alice M. Waterhouse, M. D., Medical Advisor
Walt R. Simmons, Statistical Advisor
O. K. Sagen, Ph. D., Chief, Special Studies
Philip S. Lawrence, Sc. D., Chief, Household Survey Analysis
Margery R. Cunningham, Staff Assistant

The U. S. National Health Survey is a continuing program under which the Public Health Service makes studies to determine the extent of illness and disability in the population of the United States and to gather related information. It is authorized by Public Law 652, 84th Congress.

## CO-OPERATION OF THE BUREAU OF THE CENSUS

Under the legislation establishing the National Health Survey, the Public Health Service is authorized to use, insofar as possible, the services or facilities of other Federal, State, or private agencies. For the national household survey the Bureau of the Census designed and selected the sample, conducted the household interviews, and processed the data in accordance with specifications established by the Public Health Service.

---

Public Health Service Publication No. 584-A2

# PREFACE

This report presents a description of the initial statistical design of the continuing Health Household-Interview Survey, which is a major phase of the program of the U. S. National Health Survey. The design described in this report is that used during the period, July-December 1957, which, with minor modifications, will be used throughout 1958. Except for such modifications, the design is, therefore, the basis of the statistical reports being published from the household interviews conducted during this period.

General requirements for the survey design were prepared by the Public Health Service and on the basis of these, the theoretical and operating plan of the sample was prepared by the staff of the Census Bureau. Although there are some important differences, the sample plan for this health survey draws heavily from designs previously developed by the Bureau of the Census for its Current Population Survey.

In addition to its function as the principal designer of the survey sample plan, the Census Bureau also conducts the field interviewing, and processes the data in accordance with specifications provided by the Public Health Service. Tabulation is handled on the Census Bureau's electronic computers. Final tables and published reports are planned and prepared by the Public Health Service.

Principal responsibility for development of the statistical design and preparation of the text of this report was shared by William N. Hurwitz, Harold Nisselson, Walt R. Simmons, Joseph Steinberg, Joseph Waksberg, and Theodore D. Woolsey. (Messrs. Simmons and Woolsey are members of the U. S. National Health Survey staff; Messrs. Hurwitz, Nisselson, Steinberg, and Waksberg are staff members of the Bureau of the Census.) They were assisted by numerous members of the Census Bureau staff, including especially Katherine G. Capt, Robert H. Finch, Jr., Mary J. Jaracz, Garrie J. Losee, and Helen M. Lucas.

# CONTENTS

STATISTICAL DESIGN OF THE

# HEALTH HOUSEHOLD-INTERVIEW SURVEY

## 1. INTRODUCTION

The program of the U. S. National Health Survey is a statistical measurement of the extent of illness, disability, and related conditions of the population. This program consists of several distinct but related parts. One of these is the collection of data on health through a continuing Health Household-Interview Survey. A second main part of the program is a series of surveys which utilize procedures other than household interview as the source of data on health. A third phase of the program evaluates procedures and results and develops improved techniques of measurement.

The present report describes the statistical design of the Health Household-Interview Survey. In addition to setting forth the pattern of the Survey as it was initiated in July 1957 and as it functioned in its first year of operation, the report will emphasize two further points. One is that the household interviews, while independent in a statistical sense of other surveys in the program, are but one very important component of the broader undertaking which is the U. S. National Health Survey (NHS). The second is that the household survey constitutes an evolutionary program which may be expected to change as experience accumulates and which, at any given time, is expected to fulfill only those objectives of the National Health Survey for which it is the most appropriate vehicle.

Substantive findings from the household-interview survey are being published by the Public Health Service in a sequence of numbered documents identified as Health Statistics, Series B. Technical reports and methodological studies are issued in Series A, and include this report on statistical design.

Arrangement of material in the present report is intended to facilitate use by two different groups of readers. It is hoped that the body of the report will be of interest to and readily readable by all professional persons concerned with health problems and those interested in research methods. Several technical appendices have been added for the benefit of statisticians, but contain material which may be informative for a wider audience.

## 2. BACKGROUND AND OBJECTIVES

### History

A detailed account was given in the first publication in this Series[1], of the background, need for, purposes, and expected product of the U. S. National Health Survey. That story will not be duplicated here. However, it may be helpful to recall very briefly a few highlights of the period which preceded initiation of the operating program in the middle of 1957.

By 1957 it had been more than 20 years since the last major survey to obtain comprehensive statistics on diseases, injuries, and impairments in the general population of the United States. Carried out in 1935-36, that survey was a major project in which 737,000 urban households were visited by interviewers to obtain data on morbidity, impairments, and health characteristics. It remains a landmark in the field.

In the years since 1936 there have been a number of community studies of morbidity, prominent among which are the names of Hagerstown and Baltimore, Md.; Pittsburgh, Pa.; Hunterdon County, N. J.; Kansas City, Mo.; New York City; and California (both San Jose and a statewide study). These

studies, as well as occasional experiments with supplements to the Census Bureau's Current Population Survey, demonstrated that the interview method is capable of providing useful information about the amount and distribution of diseases and injuries together with related information such as the accompanying loss of time from work or other usual activities.

In January 1949, the U. S. National Committee on Vital and Health Statistics was established. Subcommittees were established in December 1949 and October 1950 to study the needs for current morbidity statistics. As a result of their recommendations, a third Subcommittee was established in February 1951 under the chairmanship of Dr. W. Thurber Fales of Johns Hopkins University, and instructed to draft a "Plan for a national morbidity survey keeping in view the interests of local areas." After careful study, this Subcommittee recommended that several steps be taken, and in particular: "That a continuing national morbidity survey be conducted .... Its purpose would be to obtain data on the prevalence and incidence of disease, injuries, and impairments, on the nature and duration of the resulting disability, and on the amount and type of medical care received. The data would be obtained from a probability sample of households" (page 28 of reference 1).

## Public Law 652 and NHS Objectives

In the summer of 1955, the Department of Health, Education, and Welfare proposed legislation for a continuing health survey, closely paralleling recommendations of the Subcommittee. The proposal was included in the President's recommendations on health matters, received bipartisan support in Congress, was enacted into Public Law 652, 84th Congress, and was signed by the President on July 3, 1956. Later the same month appropriations were made available for planning and pretesting during the fiscal year ending June 30, 1957.

The law authorizes the Surgeon General of the Public Health Service to make continuing surveys and special studies of the population of the United States to determine the extent of illness and disability and related information such as: the number, age, sex, ability to work or engage in other activities, and occupation or activities of persons afflicted with chronic or other disease or injury or handicapping condition; the type of disease or injury or handicapping condition of each person so afflicted; the length of time that each such person has been prevented from carrying on his occupation or activities; the amounts and types of services received for or because of such conditions; and the economic and other impacts of such conditions.

A significant feature of Public Law 652 is that it not only provides that substantive data be assembled, but in addition, directs the Public Health Service, "to develop and test new or improved methods for obtaining current data on illness and disability and related information."

Thus legislative intent looks to the establishment of health statistics as noted in the law, and foresees "... continuing surveys ... special studies ... [and] develop [ing] and test [ing] new and improved methods" as the objectives of the U. S. National Health Survey.

## Planning and Pretesting the Household Interviews

Throughout the fiscal year ending in June 1957 plans were developed for organizing and carrying out the household survey which had been contemplated by the Subcommittee and authorized by Congress. The law contained the provision whereby the program could secure the assistance of other Federal agencies, as well as private persons or agencies, in carrying out its responsibilities. Under this provision, the NHS made arrangements to utilize the very extensive resources and experience of the Bureau of the Census in planning and conducting the household-interview survey.

From the beginning, it was clear that the National Health Survey should be a general multipurpose undertaking, rather than a study with some single specific limited objective. This concept meant that presurvey planning was particularly important. It required a careful review of previous efforts, a weighing and evaluating of a large number of possible alternatives, so that the new survey might be sufficiently comprehensive to cover many of the desired objectives, while at the same time not to be so diluted as to deal inadequately with all topics.

By February 1957, general structure of the survey had been determined, samples had been drawn, and a tentative questionnaire and field instructions had been drafted. A pretest of 1,200 households was conducted in Charlotte, N. C., to provide a complete field trial of procedure. The pretest was used also for training field supervisors for the national program. The next month was devoted to polishing the questionnaire and procedures, and to hiring and training interviewers. In the 2 months of May and June, the entire nationwide organization went through a shakedown and training period with interviewing and editing proceeding just as though the survey were in operation. Official collection of data began the first week in July 1957.

# 3. SUMMARY OF STRUCTURE OF HEALTH HOUSEHOLD-INTERVIEW SURVEY

## Role of Interview Survey

As noted in the previous section, the program of the U. S. National Health Survey is intended to be an intensive and sustained undertaking to provide morbidity and health statistics, utilizing whatever resources and methods are appropriate to the task. The program is expected further to evaluate existing sources and methods and to develop new methodologies.

Among possible sources of data a prominent position goes to medical and health records. These include such originating places as hospitals, physicians' and dentists' offices, and insurance records of several kinds. They include, too, reporting under governmental regulation of certain types of morbidity and mortality, and especially the filing of death certificates.

Another potentially significant source of information may lie in samples of persons who are given clinical tests and measurements or general health or medical examinations.

All these sources, and others, are to be explored by the NHS. Several pilot projects in these areas already have been initiated.

However, a considerable body of opinion considers the household interview as one of the most promising sources of data on health.

There are limitations to the accuracy of diagnostic and other information collected in household interviews. For diagnostic information the household respondent, can, at best, pass on to the interviewer only the information the physician has given to the family. For conditions not medically attended, diagnostic information is often no more than a description of symptoms. However, other types of facts, such as those concerning the circumstances and consequences of illness or injury and the resulting action taken or sought by the individual, can be obtained more accurately from household members than from any other source since only the persons concerned are in a position to report all of this type of information. Furthermore this type of survey facilitates greatly comparison of the ill population and the well population, and assessment of relative impacts of a variety of illnesses and impairments. The Health Household-Interview Survey described in this report is the vehicle being used by the U. S. National Health Survey to produce data presently believed to be most appropriately obtained from members of the household.

## Evolutionary Pattern

Continuity and comparability of estimates for different time periods are desirable objectives, and will be given attention in the interview survey, especially when changes are proposed but they will not have overriding priority. A substantial portion of resources and energy of the NHS, at least during its early years, is to be devoted to studies and evaluation of quality of data input, to efficiency of collection and processing, and to usefulness of output. It is expected that these activities, augmented by the active and constructive criticism of users, will lead to a program which is changing in response to need in scope, content, method, and specific product.

Although the interview survey has only had 1 full year of operation, already changes have been made in sample design, questionnaire, and collection and processing procedures. The description given in the following pages is in all major respects that which was in effect through the first year of operation, although minor changes occurred from one quarter to another. Quantitative references such as sample sizes and noninterview rates apply for the most part specifically to experience in the first 2 quarters of operation.

## The Questionnaire

The questionnaire is a 9-part document which is handled by the interviewer rather than the respondent, and on which the interviewer transcribes replies of the respondent. Most replies can be recorded by checking proper boxes on the form. The text of the questionnaire is supplemented by 6 check list cards which are shown to the respondent at appropriate points in the interview. The check lists clarify certain questions so as to aid the respondent in understanding types of answers required and in recalling specific experiences.

Physically, the questionnaire is of the book type, providing separate columns for each of 7 possible members of a household. If a household contains more than 7 members, more than 1 questionnaire is used.

A facsimile of the questionnaire is contained in Appendix I.

It is planned that items on the questionnaire may be divided into 2 groups—not separately ex-

hibited in the present format. One group consists of a core of basic questions which will be retained in relatively unchanged form over an extended period of time. The second group consists of supplementary questions which will be included temporarily for blocks of 1 or a few calendar quarters. This general plan provides for the retention of regular series of basic statistics, and at the same time permits flexibility in securing occasional measures of a wider class of phenomena.

As initially used, the questionnaire carries 40 items for identification of households and persons and socioeconomic description of respondents. (A question to which the interviewer must secure an answer is interpreted as one item in this count. The same interpretation applies in the following counts.) It includes 12 general questions on the presence or absence of illness, accidents, impairments, or conditions for each member of the household, and 54 detailed questions for each person—for whom the questions are appropriate—on details of illnesses, accidents, and impairments, and on medical, dental, and hospital care. For most questions, the recall period is the previous 2 weeks. But for some items of low incidence, for which memory is reliable, such as hospitalizations, the recall extends over the year previous to the interview.

Interviewing is conducted in the home, whenever possible with the individual person if over 18 years of age, and otherwise with a responsible adult member of the family.

A separate report on the questionnaire is in preparation. It will treat more thoroughly the definitions, concepts, scope, and content of the schedule. In addition, each report issued on a substantive health topic treats that part of the questionnaire which applies most directly to the topic under study.

### Sample Design, Survey Methods, and Estimation

The sampling plan of the survey follows a highly stratified multistage probability design which permits a continuous sampling of the civilian population of the United States. The first stage of the design consists of an area sample of 372 from among about 1,900 geographically defined primary sampling units (PSU's) into which the continental United States has been divided. A PSU is a county, a group of contiguous counties, or a Standard Metropolitan Area.

With no loss in general understanding, the remaining stages—which consist of a series of samplings of successively smaller parcels of land—can be telescoped and treated at this point in the report as an ultimate stage. Within PSU's then, ultimate-stage units called segments are defined, also geographically, in such a manner that each segment contains an expected 6 households in the sample. For each week a random sample of about 120 segments is drawn. Persons in the approximately 700 households in those segments are interviewed concerning illnesses, injuries, chronic conditions, disability, and other factors related to health.

Household members interviewed each week are an independent representative sample of the population, so that samples for successive weeks can be combined into larger samples for, say, a calendar quarter or a year. Thus, the design permits both continuous measurement of characteristics of high incidence or prevalence and, through the larger consolidated samples, more detailed analysis of less common characteristics and smaller categories.

The national sample plan over a 12-month period includes approximately 115,000 persons from some 36,000 households in about 6,000 segments, with representation from every state. The design is such that tabulations can be provided from the annual sample for various geographic sections of the United States and for metropolitan, urban, and rural sectors of the Nation.

Estimation is accomplished by a technique which insures that sample results are consistent with official Census Bureau estimates of current population by age, sex, and color, and which secures significant reductions in sampling variance. Technically, this procedure is a 2-stage ratio estimation. Subsequent sections in the body of this report and in the Appendices describe leading features of the design in greater detail.

## 4. SURVEY PROCEDURE

### Collection of Data

Data are collected through a household interview. Over the Nation there are 120 interviewers, trained, directed, and guided by 17 supervisors located in Census Bureau Regional Offices. The supervisors are career Civil Service employees whose prime responsibility is the National Health Survey. They have administrative and clerical support from the Census Bureau field organization, and direct technical guidance from a Health Statistics Branch in the Washington office of the Census Bureau.

The interviewers (initially all women) are part-time employees, selected through an examination and testing process which is administered by the supervisors, according to specifications set in Washington. The amount of work done by an interviewer varies depending on density of the sample near her home location. A typical interviewer may have 26 assignments in a year, or an average of 1 assignment each 2 weeks. Usually an assignment consists of interviews in approximately 12 households. Including training, travel, and call backs, the typical interviewer is employed an average of 12 hours per week.

Training for both supervisors and interviewers is a process for improving and controlling the interview and data from it. As such, it is a procedure, parts of which must continue throughout the life of the survey, and is not an activity which could be completed at the beginning of the operations.

The supervisor is given 5 kinds of training beyond the Civil Service requirements for initial appointment to the job.

First, the supervisor is supplied with written background materials setting forth the history, objectives, and purposes of the undertaking. Similarly he is given detailed instructions covering every aspect and item of field operations. He studies the materials, does practice exercises, and takes written examinations.

The second block of training, for the first group of supervisors, was participation for 2 weeks in the dress-rehearsal pretest of the survey which took place in Charlotte. Replacements have similar experiences while serving as understudies to another supervisor.

The third type of training comes from the continuing flow of written instructions and correspondence, and of evaluations of performance sent out from Washington. The latter come from quality-control and quality-checking operations performed in Washington as part of the editing processes.

Twice a year (3 times the first year) supervisors over the Nation are assembled for a 2-day review of program objectives, new developments, and selected procedural problems. These sessions permit, of course, a helpful exchange of ideas among supervisors and between the field supervisors and the Washington staff.

Finally, the supervisor has the advantage of continuing experience since his regular job includes the training of interviewers, observation of interviewing for new interviewers, and personally reinterviewing a subsample of households as a part of the quality-control program.

As stated above, the prospective interviewer is selected through a process of a written examination and testing of general intelligence and for aptitude for survey operations which she would be expected to perform. The new interviewer is then given a 5-day initial course of training. This course consists of 5 types of activity: (1) Instruction from a field supervisor on purpose and general characteristics of the survey. (2) A detailed page-by-page study of all interviewing instructions in which interviewer and supervisor go through all instructional material together. (3) Classroom Practice Exercises, in which the interviewer solves written problems and with the supervisor subsequently determines correct answers—these are exercises rather than tests, and only the interviewer knows definitely how well she has succeeded. (4) Home assignments which also are written answers to problems, which are treated more in the nature of tests and in which results are discussed by interviewer and supervisor. (5) Practice interviewing in households under direct personal observation by the supervisor. The study of instructions, the practice exercises, and the home assignments are distributed throughout the 5-day period.

If the prospective interviewer successfully completes the training course, she begins operational interviewing, her first assignment being carried out again under direct personal observation by the supervisor.

After approximately 1 month, a new interviewer is given further Home Assignments which again are graded and discussed, if necessary, by the supervisor. Subsequently, in common with all interviewers, she spends 2 hours each month on such assignments.

Each quarter the supervisor recontacts about one sixth of the households in his part of the sample. He audits the household information obtained earlier and reinterviews independently one predesignated member of the household. He compares differences between the two interviews and attempts to determine which information is correct. These reinterviews are randomly distributed among the interviewers under his supervision so that control charts based on about 5 percent of an interviewer's work can be maintained. Each week, as a part of the editing process in Washington, error rates are calculated separately for each interviewer's work. These are transmitted to the appropriate supervisor for his use in further training and in tightening control over the interview process.

Two or three times each year, groups of interviewers are assembled at Regional Offices for 1- or 2-day refresher courses on objectives, methods, procedures, and special features of the survey.

After a household has been selected for the sample, a "Dear Friend" letter, signed by the Director, Bureau of the Census (fig. 1), is addressed and a few days before the expected interview is mailed to the household. This letter is intended to be a general introduction to the survey, to have the effect of adding official sanction to it, and to make it somewhat easier for the interviewer to secure an audience. If no precise address is known, this step is foregone.

When the interviewer arrives at the household, after a very brief introduction, she begins imme-

OFFICE OF
THE DIRECTOR
Form-NHS-600
(4-26-57)

DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS
WASHINGTON 25

Dear Friend:

The Bureau of the Census has been asked by the Public Health Service to act as its agent to carry out a survey to obtain information about illnesses, diseases and injuries among residents of this area. The survey is one part of the National Health Survey Program which Congress recently authorized because of the need for up-to-date statistics on the health of our people. Physicians, research workers, and other groups in health fields are much interested in the knowledge which will be gained from this survey.

Every month several thousand addresses are chosen to give a cross-section of the whole United States, and the people at those addresses are interviewed to obtain the necessary information. This month the address of your dwelling place is one of those chosen, and you will be visited by a Census Bureau interviewer within the next week or two. The interviewer will ask you a number of questions about the health of the members of your family, particularly about the illness and injuries you have had in recent weeks. Your cooperation in helping complete a questionnaire will be very much appreciated.

The information you give will of course be held in confidence. We have the assurance of the Public Health Service that the information will be seen only by authorized personnel of the two agencies and that nothing will be published except statistical summaries in which no individuals can be identified.

Sincerely yours,

Robert W. Burgess

Robert W. Burgess
Director
Bureau of the Census

*Figure 1. Introductory letter to prospective respondent.*

diately to ask the survey questions. Each question is asked exactly as phrased on the questionnaire. Required information for each person in the household is provided by the person himself, if he is a responsible person 18 years of age or older and at home at the time of the call; otherwise by a related person who is regarded as qualified to give accurate information. This definition of an eligible respondent is spelled out in some detail in the Interviewers' Manual. In summary, answers for children are given by a related adult; for a missing adult, by wife, parent, or adult son or daughter; or for an adult not related to the head of the household, only by himself or a related adult. Early experience indicates that for persons over 18 years of age, 58 percent are "self-respondents," while the remainder for whom another person was the informant are designated "proxy-respondents". The interview averages 40 minutes. Immediately following the interview a "Thank You" letter signed by the Surgeon General of the Public Health Service is handed to the respondent (fig. 2).

In order to minimize travel time, workloads are so arranged that when an interviewer is in a neighborhood for an interviewing assignment, he carries out necessary listing operations for segments which are in that same neighborhood and which will appear in samples for the next 2 calendar quarters. Appendix VI sets forth in some detail the manner in which assignments are randomized over each quarter so that each week's interviewing constitutes a random sample of the population, and within reasonable arrangements of workload is widely diversified by geography and interviewer.

The following statistics for the first 6 months of operation shed added light on selected aspects of the collection process. Of all addresses initially scheduled for inclusion in the sample, 14 percent had become, by time of call, what are designated as Type B or Type C exclusions, which are types of addresses which should not be interviewed: dwelling units which are demolished or which on more careful inspection were found to be outside chosen sample segments; households which were deleted in the field, according to instructions, through subsampling operations (details on this step are set forth later in the report); households which were vacant; or households whose members had residence elsewhere. Of those households in which an interview should have been conducted, 6 percent were noninterviews. One percent were refusals, and five percent were not interviewed because of all other reasons, but principally because no one was at home after repeated call backs.

In about 63 percent of households, interviewing was completed on the first visit. Percent of households for which various numbers of revisits proved to be necessary are shown in the following breakdown.

| Number of visits | Percent of all households |
|---|---|
| All cases | 100 |
| 1 | 63 |
| 2 | 24 |
| 3 | 9 |
| 4 | 3 |
| 5 or more | 1 |

## Editing and Processing

The interview is recorded initially in the book questionnaire, Form NHS-1. This form is reviewed for completeness and proper identification of person and household, but otherwise not edited by the supervisor in the Census Regional Office. Reports are batched and transmitted to the Census Bureau in Washington for editing and further processing.

In Washington, certain control operations are performed, reported information is coded with special attention being given to medical coding, and to adequacy of data for medical coding (editing reports on inadequate information are returned to Regional Offices for future use in training and interviewer control), and the data are transcribed to document-sensed cards and then to punch cards. These cards are processed on conventional punch-card equipment mainly for purposes of interviewer control and for a more thorough check for completeness of entries. Rejects are returned to clerks for review and correction. Corrections and additions are punched and added to the deck. Information on cards is then transferred to magnetic tape, and further processing is handled on Univac electronic computers.

The computer carries out 4 basic operations: (1) an edit of the raw reports; (2) the generation of data from edited reports (e. g., by counting number of chronic conditions reported for a person, to generate the statistic "number of chronic conditions reported for a person"); (3) estimation of specified statistics, including all necessary computational steps such as insertion of sampling rates and adjustment for noninterview; and (4) arrangement of estimates into derived statistical tables.

As for any job of processing and editing returns in a sizeable survey, a myriad of steps is necessary. Most of these need no mention in this account. A few circumstances are worth noting.

Information moves through 4 separate channels in processing, each channel being identified as a card, and each card containing the class of information indicated by its title. The four channels are household cards, person cards, condition cards, and hospital cards.

In nearly all surveys the choice of definitions and of categorizing devices is critical to the un-

NHS-601

# *Thank You*

. . . for the assistance you have given the Census Bureau interviewer who just visited you.

It is only through the cooperation of you and others who are being visited that a health survey such as this one can be carried on, and we thought you would like to know how the information you have given will be used.

It will, of course, be held in confidence. When combined with information given by other persons in this and other communities, it will reflect health conditions throughout the United States and provide new knowledge to improve the health of the American people. It is because such knowledge is now lacking that Congress recently authorized the National Health Survey--of which the interviewing in this area is a part.

The National Health Survey will be collecting information on other aspects of health, and it is possible that we may wish to ask for your further cooperation at some time in the future. Meanwhile, thank you for your help today.

L. E. Burney
Surgeon General, Public Health Service

*Figure 2. Letter of appreciation to respondent.*

dertaking. This statement applies with special force to the Health Household-Interview Survey. In many ways classification determines the scope of the project, decides whether data on a particular topic will result, controls the meaningfulness of chosen blocks of information, influences the presence or absence of bias in the measurement process, and generally conditions the utility of survey results. For these reasons, and because adequate treatment of this one matter is lengthy, a separate report on definitions and classifications is being prepared for issue in this series. Only a few remarks are included here.

Wherever possible, standard definitions and classifications have been employed. Thus, "dwelling unit," "household," "Standard Metropolitan Area," "family," and many other terms are defined as in the Decennial Population Census or other widely accepted operations. Similarly, the International Statistical Classification of Diseases, Injuries, and Causes of Death is the basis of classifying health conditions; demographic, social, and economic measures have been grouped into classes which conform, it is believed, to most common practice.

Classifications have been predesignated and are fixed, and the questionnaire is geared in most areas to such a system. Very little latitude is given the respondent to create new classes through his replies, but translating replies into a specific code is still a relatively difficult process, especially for the medical coding.

Accordingly, medical coding for each condition initially has been done independently by 2 coders, with differences being umpired by a coding expert. Information concerning the nature of coding difficulties is being assembled to permit partial verification and process-control techniques for the medical coding.

The use of punch-card equipment for early phases of editing was dictated at first by consideration of workloads on available equipment, and of start-stop requirements for some of the steps. Some of these operations have been shifted to the computer, and others will be later.

When reports are received in Washington, they go through a control procedure at a "control desk." This procedure, in addition to routine housekeeping checks, includes 3 operations with statistical significance.

1. In the prelisting step, an "expected" number of households in each segment was determined in the field and reported to Washington. Incoming reports must account for the same number of households or explain discrepancies. Any segment for which an unexplained discrepancy is found is reconciled through recontact with the Census Bureau Regional Office if time permits. In instances in which tabulation cutoff time prevents this, the case is later called to the attention of the Regional Office so that it will be used in initiating necessary tightening of supervisory controls on listing, interviewing, and clerical operations.

2. In some segments it will have been found, either earlier in Washington or in the field, that a chosen segment clearly contained more than 20 households. In these cases the segment was subsampled so that the final subsample contained roughly 6 households. The subsampling fraction is noted at the Washington control desk and an adjusting order is transmitted to the computer.

3. A third type of review at the control desk adjusts the sampling fraction for households and persons from special dwelling places, such as reformatories, homes for the aged, or hotels for transients.

The general purpose of all operations at the Washington control desk is to assure that data moving into the editing and tabulating stream are, with respect to coverage and weighting, in agreement with the survey design, within narrow tolerances of error.

## Evaluation and Control of Data

A substantial proportion of total resources of the household-interview survey is devoted to control, evaluation, and improvement of quality of data input.

There are 4 very broad areas of activity which have impact on quality of data, which are not discussed here in this connection, but which are parts of the U. S. National Health Survey, and which are listed by title in order to place in perspective those items which are displayed in the following paragraphs as devices for control and improvement of data. The 4 areas are (1) the over-all survey design, including concepts, definitions, and general plan of operation; (2) operating control, in the sense of maintaining general adherence to design, including proper use of training and supervision; (3) utilization of comparative analysis of data, including external checks against other sources of health information, and especially against medical and health records; and (4) provision for organized outside review and criticism of both methods and products through the creation of both governmental and nongovernmental advisory committees, and the use of consultants.

Aside, then, from these broad areas just named, there are 3 types of operation which are integral parts of survey procedure, and which are principally devices for control and improvement of quality of data.

The Reinterview Procedure.—Already mentioned, in connection with training and supervision of interviewers, is the reinterview program. The supervisor regularly recontacts about one sixth of the households in his part of the sample, and thus, about one sixth of the households assigned to each interviewer. The supervisor audits the household information previously secured by the original interviewer, and reinterviews 1 predesignated member of the household. Three main

purposes, and several lesser objectives, are served by this procedure. The first purpose is that of training and quality control for the interviewers. The second is measurement of interviewer variability. The third is detection of interviewer bias—to the extent that the more expert supervisor can discover it. This means embracing the assumption that the supervisor, using the same questionnaire and the same procedure that were used by the initial interviewer, but being more thoroughly trained, will secure data which may be considered the standard which the interviewer should have met—and possibly did. On the reinterview, all adults must be interviewed as self-respondents rather than proxy-respondents. Thus there is a component of variance from self- vs. proxy-respondent as source. This component has some diluting effect on measurement of interviewer contributions to bias and variance, but its existence may make it possible to determine the extent of bias (if any) caused by the proxy-respondent.

Processing checks and controls.—At each principal processing step, controls are established either by verification through duplicate processing, or by sample verification techniques, based primarily on process control. Thus far, error rates beyond preliminary standards are called to the attention of the responsible operating supervisor, with a recommendation that steps be taken to reduce the error rate. Further study of error rates and their probable impact on estimates are expected to lead to a better-balanced set of standards. They will lead also to greater use of sample verification and reductions in 100 percent duplication of editing steps.

Internal editing and consistency checks.—Reference was made earlier, and will be made again when the details of estimation in the survey are discussed, to editing routines designed to make questionnaires internally consistent, and to eliminate "impossible" responses. This is an area in which the number and type of possible checks are unlimited. Experience must be the guide in deciding how much editing is profitable. As implied earlier, the first objectives are to insure that data are consistent and not obviously incorrect. More penetrating edits are to be tested.

# 5. SAMPLE DESIGN

## The Multistage Design

As noted in the summary on page 4 of this report, the Health Household-Interview Survey rests on a highly stratified, constructively 2-stage probability design. Actual selection of sample units takes place in a multistage process, which is modified further by the use of 3 selection zones and 41 subuniverses. The design is termed "constructively 2-stage" because the first sampling step is the selection of 372 primary sampling units from among some 1,900 areas into which the country has been divided, while the remaining steps lead effectively to a second or ultimate sampling stage in which small segments or clusters of an expected 6 households are chosen for inclusion in the sample from within the PSU's selected in the first step.

The following paragraphs describe principal features of the design, and the manner in which the sample was drawn. Additional technical notes on selected aspects of the design are included in Appendices II through VII. In particular, algebraic statements of the estimating and variance equations are given in Appendices II and III. Still further insight on the topic can be gained from consulting Chapters 7, 8, 9, and 12 and Appendix B of reference 2, since much of the theory underlying the sample design of the health survey is set forth in this book.

Primary sampling units.—The PSU is a county, a group of contiguous counties, or a Standard Metropolitan Area. A total of 1,900 PSU's exhaust the land area of the continental United States. Formation of such PSU's is an art rather than a science, although several clear-cut principles and rules were helpful. Prominent among these are the following 4:

1. PSU's should be units for which a wide variety of descriptive statistics is available, since this permits the PSU's to be stratified or classified in an efficient manner.

2. When the PSU is used by a large surveying organization, there are distinct economies in using the same set of PSU's for more than 1 survey. Consequently there are advantages in having the PSU conform to administrative structure in the field, and in having the unit adaptable to many social and economic objectives.

3. For technical sampling reasons, the greater the internal heterogeneity of the PSU, the more efficient it is. This principle tends to produce physically large units.

4. Contrastingly, costs per ultimate sample unit (i. e., cluster of sample households) tend to increase with transportation distances between ultimate units within a PSU, and thus to increase with the size of the PSU. This factor has limited the size of a PSU to not more than a few neighboring counties.

The above principles led to formation of the 1,900 PSU's, which are also used in other Census Bureau surveys, and which, with a few exceptions, have these features: The building block or smallest structural component of the PSU is a county; each PSU in the Western United States contains a population of at least 7,500 (1950 Census), and in other parts of the country a population of at least 10,000; each western PSU contains not more than 2,000 square miles and other PSU's not more than 1,500 square miles—unless the single county is larger, which in the West resulted in many PSU's having less than 7,500 persons; and, with the qualification that each Standard Metropolitan Area is a PSU, the PSU is kept as internally contrasting as possible in socioeconomic terms.

Stratification of PSU's.—Sampling theory makes it clear that if units to be sampled can be classified into categories or strata whose members tend to be relatively alike within strata and relatively unlike between strata, and drawings made from those strata, then resulting sampling variances are reduced over those of samples drawn from an unstratified universe. The PSU's were stratified accordingly, the principal modes of stratification being geographic location, density of population, rate of population growth between 1940 and 1950, proportion of nonwhite, type of industry in predominantly urban areas, and type of farming in rural areas. The general sampling design contemplated drawing first-stage units with probability proportionate to size, with 1 PSU to be drawn from each stratum. Further, it was desired that separate estimates be obtainable readily for each of 41 subuniverses—to be further described later, but characterized often in the survey as Tab Areas. These specifications, augmented by an existing stratification of the PSU's, set up by the Census Bureau for other purposes, resulted in classification of the approximately 1,900 PSU's into 372 strata. Further description of the precise manner in which this was done is given in Appendix IV.

Drawing first-stage units.—From each of the 372 strata 1 PSU was selected for inclusion in the sample with probability proportionate to its 1950 population. This meant, for example, that a small PSU with 50,000 inhabitants in 1950 had only 1/20 as much chance of inclusion in the sample as did the larger PSU with 1 million inhabitants. These differential sampling rates were of course taken into consideration in subsequent sampling and estimating steps.

As indicated, the selection procedure and the specification that separate worksheet estimates be computed for each of the Tab Areas had influenced stratification. The Tab Areas initially specified were the 8 largest Standard Metropolitan Areas, and within each of 11 geographic sections, the 3 subsections composed of (1) smaller Standard Metropolitan Areas, (2) other urban areas, and (3) other rural areas. The sections and 8 large SMA's are shown on the map in figure 3. It should be understood that separate statistics will not be published for each of the different Tab Areas, but rather that data for Tab Areas can be consolidated in more than one way into broader categories for which reliable figures can be produced.

In some instances, efficient stratification resulted in a stratum being composed of 1 single large PSU. From such a stratum the single PSU enters the sample with certainty and is called a self-representing PSU. Each of the 8 largest SMA's and 102 other PSU's became self-representing PSU's. Each PSU drawn into the sample from a nonmetropolitan stratum was utilized later as the frame for both "other urban" and "other rural" tabulation areas. Table 12 in Appendix VIII shows the geographic distribution of both self-representing and nonself-representing PSU's.

Selection zones.—For sampling purposes and in order to reduce over-all variance, the civilian population in the United States is divided into 3 mutually exclusive classes or selection zones:

Zone A. Those persons living in common dwelling places.

Zone B. Those persons living in areas of "new housing."

Zone C. Those persons living in large special dwelling places.

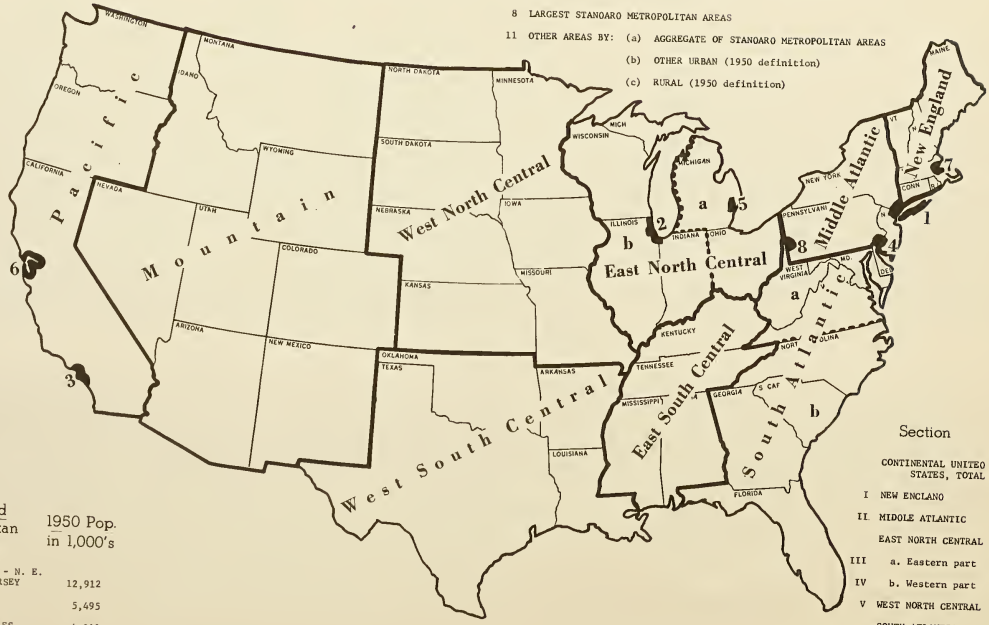
Common dwelling places include what would be ordinarily regarded as such—for example, private homes, apartment houses, and duplexes. Areas of new housing are simply those in which considerable new housing has been built since the last population census (April 1950) and which have been recorded and mapped by the Census Bureau. These may include areas which would be classed as belonging to either Zone A or Zone C except that they are positively identified as being in Zone B. Special dwelling places include such places as penitentiaries, reformatories, homes for the aged, mental hospitals, and hotels for transients.

The 372 first-stage units are identical for all 3 zones, but later-stage sampling is handled separately for each zone.

For the large special dwelling places, Zone C, lists of individual institutions and organizations in the sample PSU's were assembled from a variety of sources. These listed places are excluded from further area sampling. Special instructions for drawing samples of persons from Zone C are prepared for the different types of special dwelling places. Such persons, constituting about 2 percent of the universe, have not been included in initial tabulation of data and are not discussed further in this account.

The relationship between selection Zones A and B and between Zones B and C is slightly more complex and makes use of the principle of stratification after sampling[3] and page 468 of reference 2. One of the risks of area sampling, when using data on number of households for a prior year as the basis for selection, lies in the existence of large units of new construction built since the prior

# TABULATION AREAS: NATIONAL HEALTH SURVEY




| Standard Metropolitan Areas | 1950 Pop. in 1,000's |
|---|---|
| 1. NEW YORK - N. E. NEW JERSEY | 12,912 |
| 2. CHICAGO | 5,495 |
| 3. LOS ANGELES | 4,368 |
| 4. PHILADELPHIA | 3,671 |
| 5. DETROIT | 3,016 |
| 6. SAN FRANCISCO | 2,241 |
| 7. BOSTON | 2,370 |
| 8. PITTSBURCH | 2,213 |

| Section | | | 1950 Pop. in 1,000's |
|---|---|---|---|
| | CONTINENTAL UNITED STATES, TOTAL | | 150,697 |
| I | NEW ENGLAND | | 9,314 |
| II | MIDDLE ATLANTIC | | 30,164 |
| | EAST NORTH CENTRAL | | 30,399 |
| III | a. Eastern part | 14,318 | |
| IV | b. Western part | 16,081 | |
| V | WEST NORTH CENTRAL | | 14,061 |
| | SOUTH ATLANTIC | | 21,182 |
| VI | a. Upper part | 8,787 | |
| VII | b. Lower part | 12,395 | |
| VIII | EAST SOUTH CENTRAL | | 11,477 |
| IX | WEST SOUTH CENTRAL | | 14,538 |
| X | MOUNTAIN | | 5,075 |
| XI | PACIFIC | | 14,487 |

Figure 3.

year. This phenomenon causes no bias in estimates, but unless corrective action is taken will increase variance.

From the National Housing Inventory of 1956 the Census Bureau had available a record of large new contruction activities in many of the 372 sample PSU's. Consider these PSU's as being stratified into 2 classes: (1) those PSU's which contain areas of new housing so identified by the Inventory; and (2) all other PSU's. Segments are selected from within all sample PSU's to represent Zones A and C. Zones A and C are mutually exclusive; that is, they do not overlap. The segments selected from the class 1 PSU's are then examined to see if they fall into areas classified as new construction areas according to the Housing Inventory. Those segments from class 1 which do not fall into new construction areas and all segments selected from class 2 PSU's are retained and become the independent samples for Zones A and C. The segments or parts of segments which are contained in the areas of new construction, and which were initially drawn into the sample, are at this point deleted from the original sample. An independent sample is then taken from among the new construction areas of Zone B at the same sampling rate as Zone A. Over-all sampling ratios for all 3 selection Zones A, B, and C are identical within each Tab Area. Approximately 8 percent of the population and of the sample are accounted for by Zone B.

Selection of segments in Zone A.—Thus by far the major part (90 percent) of the sample is found in Zone A. For many purposes, it is convenient to think of the sample as consisting only of Zone A. An outline is given here of the way in which sampling within PSU's is carried out for selection Zone A. An example of the process is given in Appendix VI.

The ultimate sampling unit within the PSU is called a segment. It is a geographically defined parcel which contains an expected 6 households. Segments to be included in the sample are chosen separately for each Tab Area in a series of steps or stages.

Survey specifications resulted in a requirement that over a period of a year 144 segments are to be surveyed in each Tab Area. Within chosen segments, all households are interviewed. (As noted in Section 4, if it develops that a selected segment contains obviously more than 20 households, it is subsampled and approximately 6 households in it are interviewed.)

The selection procedure allocates the number of segments to be interviewed to first-stage units in the Tab Area in proportion to the size of the stratum they represent. Segments are drawn within PSU's through a sequence of selection of successively smaller units of area until finally a unit containing the expected 6 households is secured. This becomes the ultimate sampling unit. An illustration of the procedure is given in Appendix VI.

Samples for the year, quarter, and week.—Initial sampling is carried on in a way which makes the segments reported for each calendar quarter an independent sample of the land area of the United States. The quarterly samples are additive and thus the annual sample is 4 times the size of the quarterly samples. The samples are also randomized by weeks within each quarter, so that each week's interviews become a random sample of the population and the weekly samples are additive within the quarter. The detail by which this is accomplished is illustrated in Appendix VII. The full survey design is effective over each quarter. The weekly samples are unbiased but necessarily follow a more restricted design, on the average depending upon a first-stage selection of 60 rather than 372 PSU's.

## Mapping of Segments and Listing of Households

For each segment in the sample, the interviewer is furnished 2 maps: a Key Map and a Segment Map.

The Key Map shows the general location of the segment and may be a county highway map or a city street or block map. The segment number and approximate location of the segment (shown by the large dot beneath "Hillcrest Avenue" in figure 4 are entered on the Key Map.

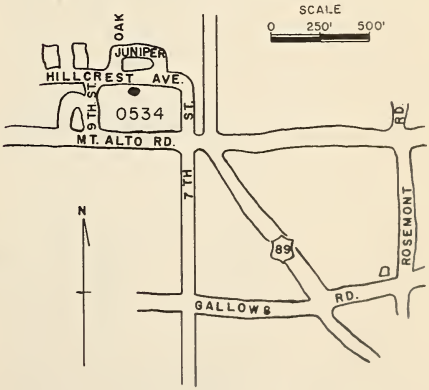



*Figure 4. Key Map showing Segment 0534.*

The Segment Map shows exact boundaries of the segment and may show either the exact or general location of some structures within the segment, depending on the kind of map available. In some cases no structures are shown on the map. Two illustrations of Segment Maps are shown in figures 5 and 6. The segment boundaries, in any case, are outlined on the map.

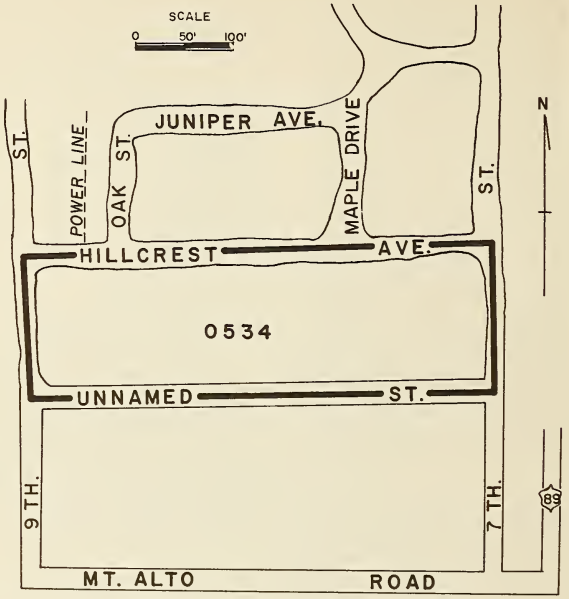



*Figure 5. One type of Segment Map.*

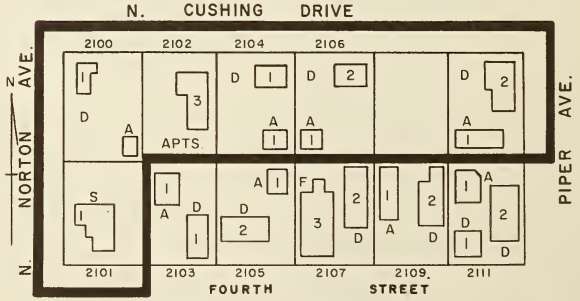



*Figure 6. A second type of Segment Map.*

*D - Dwelling F - Flat S - Store Apts. - Apartments A - Garage*
*The numbers inside indicate the number of floors in the structure and the numbers along the margin are street numbers.*

Interviewers are instructed to list all households or dwelling units in the selected segments. The meaning of this instruction is summarized in the sentence: "Write on prepared forms the addresses or other descriptions of all places where people live or might live, including such places as ordinary house dwellings, apartments, duplexes, trailers, tents, houseboats, converted boxcars, and rented rooms, including everything which lies inside the defined segment." The instruction is supported and amplified by the maps and a 93-page indexed listing manual. The listing operation is conducted at a time prior to interviewing, thus providing 2 checks on coverage: one at the time of listing and a second at the time of interviewing.

Summary of units.—Several different kinds of units and categories are mentioned in this section and in the Appendices. It may be helpful to recapitulate in capsule form the principal elements of terminology.

| | |
|---|---|
| Tab Area | - One of 41 subuniverses, defined by geographic boundaries and by size and density of population. |
| PSU | - Primary sampling units consisting of 1 or a group of contiguous counties: about 1,900 of them in the United States; 372 in the NHS sample. |
| Strata | - 372 socioeconomic classes into which the PSU's are grouped. |
| ED | - Enumeration District, a geographic subdivision of a PSU, usually containing between 50 and 1,000 households. |
| Segment | - A subdivision of an ED containing an expected 6 households. This is normally the ultimate sampling unit in the survey. |
| Selection Zones | - Strata in a different dimension, based upon type of dwelling unit, and utilized in reducing variance. |
| Dwelling Unit | - Place where persons live or might live. This is the unit listed for subsequent interviewing purposes. |
| Elementary Unit | - There are 4 elementary units or channels for processing information which are utilized in the survey: (1) the household; (2) the person; (3) the health condition—illness, injury, chronic condition, or impairment; and (4) the episode of hospitalization. |

## The Estimating Process

Some aspects of the estimating process were treated in Section 4 under Editing and Processing, and other aspects are influenced, of course, by the sample design which has just been discussed. In what follows in the present section, the focus of attention is on the estimating problem as such.

The estimation process in the health survey is basically simple, although actual procedure includes a considerable number of steps. Leading reasons for the apparent complexity are 4 in number, growing largely out of the fact that the survey produces a variety of estimates in several dimensions:

Geographic scope.—The survey yields worksheet figures for the Nation as a whole, and also for constituent Tab Areas. The Tab Areas can be combined into geographic divisions of the country, or into classes which reflect size and density of the population in the community.

Type of statistic.—Three variations may be distinguished under this heading. (1) The number or proportion of persons in the population with a specified characteristic, such as having 1 or more chronic conditions, or not having visited a physician within the year immediately previous to the week of interview. (2) Estimated volumes of events arising from tabulating answers to such direct questions as, "How many days were you in the hospital, not counting the day you left?", with editing converting the reply to "Number of hospital days in past year." (3) The incidence of a particular disease or health condition, built up from cumulating occurrences over 2-week periods as reported by persons interviewed in successive weeks.

The first type of statistic named above will be recognized as an instance of binomial estimation (modified of course by the structure of the sample design), since each individual respondent will either have or not have the specified characteristic. The second type of statistic is like the first except that the population measures involved are quantitative rather than qualitative variables, and consequently estimation is not binomial. The distinction between the second and third types of statistic is sharpened perhaps with an example. Approximately 115,000 persons are interviewed each year in the health survey, about 2,200 each week. Each of the 115,000 persons, in effect, gives the interviewer the number of days he spent in the hospital in the previous year, and thus provides data which permit an estimate of the number of days of hospitalization experienced by living persons in the year previous to the week of the interview. This is a type (2) estimate. Similarly, each week approximately 2,200 persons report their days of hospitalization in the previous 2 weeks. Summing these reports over 52 weeks of interviewing and taking account of the 1-week overlap in reference periods for adjacent weeks of interviewing would provide the basis for a second estimate of a year's hospitalization, this time the resulting statistic being of type (3). More is said later on the procedure whereby estimates of type (3) are produced.

Some might also wish to distinguish, under this title, between estimates of an aggregate, such as total number of physician visits for a specified

class of persons, and estimates of a rate which would express the number of physician visits for the class per 100 persons in the class.

Time reference.—Since the sample is continuous, it can be used to provide estimates based on interviews of the population over a week, a quarter, a year, or other time intervals. Also reference periods for occurrence or volume of events can be varied widely, ranging for some items from a week to any multiple of weeks within the history of the survey.

Form of estimate.—The statistics produced from the stratified design through 2 stages of ratio estimation are the products of a design which is much more efficient than a simple random sample would have been, but which necessarily require somewhat more elaborate computation.

Steps in estimation.—In the interest of bringing out main threads of the estimating story, obscured as little as possible by the crosscurrents just noted, the remainder of this section is written mostly around the production of estimates of an average number of persons with a specified characteristic, the average being based on interviewing over 13 weeks. The population referred to is the civilian noninstitutional population of the continental United States rather than that of one of the Tab Areas. An aggregate rather than a proportion or rate is the statistic under observation. Occasional variations from this pattern will be necessary.

Step 1

As indicated earlier, incoming reports are passed through controls to insure that the data input to the computers is consistent with sample design, properly coded, and capable of being tabulated.

Step 2

A series of mechanical edits are carried out on the computers. These edits make the questionnaire internally consistent, and adjust or account for item nonresponse.

Step 3

Into each record of an elementary unit (person, household, condition, and hospitalization) basic sampling inflation factors are inserted. This step takes account of all stages of sampling. The factor is the reciprocal of the combined sampling fraction which for a quarterly tabulation varies among Tab Areas from about 1-in-2,000 persons to about 1-in-19,000 persons. [Sampling fractions for annual samples are one quarter of these numbers.]

Step 4

Statistical theory demonstrates that a "ratio estimate" for any statistic is superior to an ordinary "inflation estimate" if there is correlation between the numerator and the denominator of the ratio. Specifically, if $Y'$ and $X'$ are ordinary inflation estimates of 2 characteristics of a population, Y and X, respectively, and if the "true" total X is known independently, then the ratio estimate $Y'' = \frac{Y'}{X'} X$ is a better estimate of Y than is $Y'$, provided there is correlation between $Y'$ and $X'$. In this form of estimate, the quantity $\frac{X}{X'}$ becomes a calibration factor for the survey.

This principle is utilized at 2 stages in the NHS. In the first instance it is used to reduce sampling variance between PSU's. Estimates of the 1950 population which would have been obtained from a complete enumeration of the 372 PSU's but not other PSU's in the country were compared with official 1950 population counts for each of 120 color-residence classes. Resulting factors are shown in table A.

In calculation, these factors are used in the following manner, the arithmetic being carried out automatically by the computers. Consider a sample record for a person who is white and who comes from an urban nonself-representing Standard Metropolitan Area in Geographic Region 1. All sample records for this person are multiplied by the factor 1.075380. (See 1st line, 2d column of table A.) This brings the sample data into closer conformity with population controls for the universe, introduces only trivial, if any, bias into the estimate, and reduces sampling variance.

NOTE: Steps 1 through 4 are carried out weekly, and provide a "deck of cards" (Univac tape) of edited and adjusted sample data for each week of the 13 weeks of the quarter. The "scale" of data at this point is therefore 1/13th of universe totals. Weekly data are merged later into quarterly totals. Steps 5 and later apply to the merged quarterly decks.

Step 5

Despite intensive follow-up efforts, reports on some households in the sample have not been received at the tabulation cutoff. In the first 2 quarters of operation the noninterview rate was 6 percent—1 percent refusal, and the rest for all other reasons, such as no one at home after repeated call backs. For a sample household for which no interview is obtained, any estimating procedure must necessarily impute values for each statistic for which measurement had been intended.

Adjustment for noninterviews in the health survey is accomplished by a calculation which assumes that respondents within a particular segment for a quarter represent the nonrespondents in that segment. In the rare instance in which less than half of a segment is interviewed, the noninterview adjustment is modified by evidence from

Table A. First-stage ratio estimate factors for nonself-representing[1] PSU's by residence, color, and section

| Type | | Urban | | Rural nonfarm | | Rural farm | |
|---|---|---|---|---|---|---|---|
| | | White | Nonwhite | White | Nonwhite | White | Nonwhite |
| Nonself-representing SMA's geographic section | 1----- | 1.075380 | 1.753515 | .678533 | .673733 | .579098 | .488372 |
| | 2----- | .973243 | .809129 | 1.048442 | .927872 | 1.345792 | .721580 |
| | 3----- | .755966 | .724533 | .664274 | .633703 | .723622 | .695719 |
| | 4----- | 1.328674 | .927637 | 1.158533 | 1.158533 | 1.580769 | 1.580769 |
| | 5----- | 1.210693 | .509411 | 1.461507 | .428340 | 1.640072 | .383598 |
| | 6----- | none | | | | | |
| | 7----- | .973222 | .869424 | .889450 | .715170 | .870792 | .745692 |
| | 8----- | 1.076720 | 1.027738 | .899394 | 1.349170 | .776500 | .991501 |
| | 9----- | 1.179743 | 1.179743 | 1.094866 | 1.874840 | 1.250069 | 1.250069 |
| | 10----- | .873733 | .492896 | 2.502347 | 2.502347 | .873241 | .873241 |
| | 11----- | none | | | | | |
| Nonself-representing non-SMA's geographic section | 1----- | 1.000096 | .791543 | 1.066578 | 1.084175 | .991135 | 1.003409 |
| | 2----- | 1.050026 | .891068 | .959741 | 1.243901 | .991556 | 1.301490 |
| | 3----- | 1.101374 | 1.476958 | .962977 | .883253 | .919709 | .728575 |
| | 4----- | 1.022545 | 1.142936 | .953182 | .912743 | .997690 | .871958 |
| | 5----- | 1.109184 | 1.649665 | .873888 | .873888 | 1.000277 | 1.673347 |
| | 6----- | 1.000335 | 1.065752 | 1.011195 | 1.276623 | 1.013374 | 1.156285 |
| | 7----- | .988710 | 1.044770 | 1.086449 | 1.027072 | 1.008986 | 1.071351 |
| | 8----- | .984943 | 1.028117 | 1.016688 | 1.026288 | 1.004002 | .968411 |
| | 9----- | .980173 | .977126 | 1.005968 | 1.013235 | 1.016431 | .990630 |
| | 10----- | 1.043498 | .979703 | 1.005739 | 1.108361 | .890752 | 1.199575 |
| | 11----- | 1.020053 | 1.203531 | .978919 | .807592 | 1.006944 | 1.009057 |

[1]First-stage ratio estimate factors for each of 8 large separate tabulation areas and for the self-representing PSU's is 1.000000.

reports over the entire Tab Area. An illustration of the process is given for a hypothetical Tab Area:

| Segment number | Households scheduled for interview | Households not interviewed | Segment adjustment factor | Excess noninterviews |
|---|---|---|---|---|
| 1 | 6 | 0 | 1.0000 | 0 |
| 2 | 6 | 1 | 1.2000 | 0 |
| 3 | 8 | 0 | 1.0000 | 0 |
| 4 | 4 | 3 | 2.0000 | 2 |
| ' | ' | ' | ' | ' |
| ' | ' | ' | ' | ' |
| ' | ' | ' | ' | ' |
| Tab Area total | 220 | 10 | - | 2 |

Data for the 5 reports in segment 2 are multiplied by the factor 1.2000 so that the 5 reports represent the 6 households intended for interview in the segment. Segment 4 in the example is of the unusual type (where less than half the households in the segment were interviewed) which leads to a further adjustment at the Tab Area level after a preliminary one has been made at the segment level. The Tab Area adjustment factor is the ratio of total households scheduled for interview to total households scheduled for interview less the "excess" noninterviews; that is, the factor is 220/218 or 1.0092, in the example. Data for all reporting households in the Tab Area are multiplied by this factor to account for the 2-household "excess" of noninterviews.

Step 6

Advantages of the ratio-estimating process are exploited further by the introduction of a second calibrating or ratio factor which brings the

estimates of the U. S. population derived from the health survey into agreement with independently determined controls for 76 age-sex-color classes of the population. For the first full quarter of operation these factors ranged from 0.61 to 1.36, with NHS estimates for 62 of the 76 classes coming within 12 percent of the controls. The over-all NHS estimate of the U. S. population before this final adjustment was within 0.3 of 1 percent of the control on the population.

The effect of these 6 steps is (1) to use the household survey as an instrument for obtaining percent distributions of the population by specified characteristics of illness and health conditions, and (2) to produce estimates of total numbers of persons in the population with these specified characteristics by multiplying the derived percent distribution by population controls. Rates are calculated by obtaining ratios of the appropriate estimated aggregates.

Tabulations of items other than average number of persons with specified characteristics over a quarter are obtained in a similar manner, but with variations in procedure, the particular variations depending on the nature of the item. Two examples may suggest the kinds of variations which are needed.

Consider again the type (3) estimate discussed above, in which the objective is to obtain an estimate of the total number of days of hospitalization over a year, and consider first an estimate over 1 quarter. An item on the questionnaire asks each person for the number of such days in the 2-week period immediately preceding the calendar week of interview. Each week's interviewing, since it is an independent sample of the population, produces, by the process described in the 6 steps above, an estimate of 1/13th of the total hospital days over a 2-week period. (It will be recalled that the sampling fractions have been expressed in terms of 13 weeks of interviewing, and the weighting factors have been set accordingly in the computer.) Multiplication by 6.5 yields 1/13th of the total visits for a 13-week period. Summation of samples over the quarter yields the estimate for a 13-week period. The particular 13-week period is the one extending from the 12th week of the quarter preceding the quarter of interviewing through the 11th week of the quarter of interviewing, since tabulation is geared to weeks of interviewing which lie in the calendar quarter. While this period does not correspond exactly with the 13 calendar weeks of the quarter, the displacement is small, and estimates made in this manner are used as estimates for the calendar quarter. Similarly produced estimates summed for 4 successive quarters would yield an approximate estimate of hospital days for the population over the year. This estimate does not include hospital days for persons who died within the 2-week period immediately preceding the week of interview, since the scope of the household survey is the living population in the week of interview.

A second illustration relates to combining estimates for more than 1 quarter, when the quarterly estimates have been expressed as rates. The problem might be formulated in many ways. One will suffice here. From each quarter's sample an estimate of the average number of persons who have experienced 1 or more days of bed-disability in a 2-week period can be produced. This figure divided by the average population for the quarter yields a rate. An annual rate based on experience for a year rather than for a quarter could be formed in more than one way. An acceptable solution is a weighted average rate calculated as indicated:

Let

$B_i$ be the number of persons in $i^{th}$ quarter with 1 or more days of bed-disability in a 2-week period, as estimated in the first example above,

$N_i$ be average population in $i^{th}$ quarter, and

$R_i$ equal to $B_i/N_i$ be the quarterly rate;

then the annual rate, R, may be estimated as

$$R \text{ equal to } \frac{\sum_{i=1}^{4} R_i N_i}{\sum_{i=1}^{4} N_i} .$$

## Sampling and Measurement Errors

Reliability of statistical surveys.—All statistical surveys, whether based on samples or attempted complete enumerations, are subject to potential inaccuracies. These risks include, among others, errors in conceptual formulation, ambiguities in definition and in the questionnaire, faulty classification, interviewer variability and bias, respondent bias and variability, biases from nonresponse or incomplete coverage, mistakes in editing, and tabulation errors. This broad group of imperfections can be subsumed by the term "measurement error," which includes all nonsampling hazards. Measurement error plus sampling error may be called total survey error.

Ideally it is desirable to detect all major components of total survey error, quantify each of them, and allocate resources in such a fashion that total survey error is minimized. Occasionally it is preferable to exclude from consideration certain specified components, even if they are large,

if their presence can have little impact on decisions which will be based on results of the survey—for example, it may be well to tolerate certain kinds of constant bias, if the survey is to be used principally to assess change from one point in time to another.

Measurement error.—A rather substantial portion of the total budget for the National Health Household-Interview Survey is earmarked for the study of measurement error and the evaluation of results. This topic is not covered in any detail in the present report. As noted earlier, however, the initial program includes 3 main areas of exploration: (1) built-in tests and controls, such as the reinterview operation, which will provide data on interviewer variation and bias, and as consistency controls on medical coding; (2) external special statistical checks, such as ad hoc studies of selected medical and health records; and (3) comparative analysis of data from the household-interview survey and of evidence from other sources of health information.

Sampling error.—Since estimates from the health survey are based on a sample of households rather than on a complete census of persons in the United States, they will differ somewhat from figures which would be obtained from a complete enumeration using the same schedules, instructions, interviewers, and procedures. Inasmuch as it is possible in the sample to use better trained interviewers, and in general to maintain tighter operational control than would be feasible in an attempted complete enumeration of many characteristics of a population of more than 170 million persons, it is entirely likely that the sample results are subject to a smaller measurement error than would be those from a census. The usual yardstick of sampling variability is the standard error, or the relative standard error. Appendix III sets forth the method by which standard errors for statistics from the survey are computed. The method used reflects both the chance error that arises from sampling, and a part of the variation which resides in the measurement process. It does not include the part of measurement variation which is unaffected by sample size, nor does it include any biases which may lie in the data.

For probability samples of the type of the health-interview survey, sampling reliability for any statistic from the survey can be stated roughly in these terms: A census would produce figures within 1 standard error of the published sample estimate for about 2 out of 3 of the statistics shown and within 2 standard errors of the published sample estimate for roughly 19 out of 20 of the independent statistics shown. A somewhat more precise statement might read: "In a complete enumeration conducted under identical circumstances the measured statistic would lie in the interval formed by the published sample figure plus or minus k times the standard error"; and the probability that this is a true statement is given in the following table.

| If k is, | then the statement is true approximately |
|---|---|
| 1 | 2 times out of 3 |
| 2 | 19 times out of 20 |
| 2½ | 99 times out of 100 |

Reports published by the health survey include statements of sampling reliability for principal estimates included in the report. In addition, as experience is gained, it is expected that general guides and rules of thumb will be developed whereby users of the statistics can secure approximate sampling errors for other figures, with a minimum of effort.

It may be useful to note relative magnitudes among some of the different classes of statistics which will come from the household survey.

If V is the relative standard error for a statistic which refers to an estimate for a U. S. total, then relative standard errors for the same statistic when it refers to other subdivisions of the United States usually will be of the general magnitude indicated in table B.

Table B. Magnitudes of statistics for several types of area

| Area | Rough magnitude of relative sampling standard error |
|---|---|
| U. S. total | V |
| A geographic section (e. g., New England) | 3.3 V |
| Rural United States | 2.0 V |
| The non-metropolitan urban sector of the United States | 2.0 V |
| Metropolitan United States | 1.5 V |

Similarly, if A is a relative standard error for a statistic which rests on data for a year's interviewing, the magnitude of the corresponding relative error for the statistic based on 1 quarter's sample will be about 1.7 A.

If B is a relative error for a characteristic possessed by 1 percent of the population, the relative error for a statistic possessed by 10 percent of the population will have magnitude approximately

30 percent of B; the relative error for a statistic possessed by 50 percent of the population will have magnitude of the order of 10 percent of B.

Standard errors of differences between estimates of the same statistics for 2 points in time will be 40 percent larger than the standard error of the statistic at a fixed point in time.

Finally, the reliability of an estimated rate or percent, computed by using sample data for both numerator and denominator, depends upon the size of the rate and the size of the total upon which the rate is based. Estimated rates are relatively more reliable than corresponding absolute estimates of the numerator of the rate, particularly if the rate is high. However, ratios of estimated aggregates to total population for an age-sex-color class have the same relative sampling variance as the estimated aggregate, as a result of the ratio estimating technique which was employed.

Illustrative sampling errors.—Relative sampling errors have been calculated for a number of estimated national statistics based on data for the first 13 weeks of interviewing. The extent to which these values prove to be typical must await the evidence of later data. Illustrative errors are presented in table C.

Table C. Illustrative relative sampling errors for national statistics from the U. S. National Health Survey, based on data from interviewing during the 13-week period ending September 29, 1957

| Statistic | Size of statistic (000,000) | Relative standard error |
|---|---|---|
| Number of bed-days for medically attended chronic conditions in last 12 months | 756 | 0.010 |
| Number of visits to the doctor | 199 | 0.022 |
| Number of acute conditions | 70 | 0.030 |
| Number of acute conditions, medically attended | 47 | 0.042 |
| Number of persons with chronic limitation of activity | 17 | 0.030 |
| Number of persons injured in accidents | 14 | 0.051 |
| Number of persons injured in motor-vehicle accidents | 1 | 0.175 |

# APPENDIX I

## ILLUSTRATION OF CONTENT OF INITIAL BASIC HOUSEHOLD QUESTIONNAIRE

The items below show the exact content and wording of the questionnaire used in the household survey. The actual questionnaire is designed for a household as a unit and includes additional spaces for reports on more than one person.

CONFIDENTIAL: The National Health Survey is authorized by Public Law 652 of the 84th Congress (70 Stat 489; 42 U.S.C. 305). All information which would permit identification of the individual will be held strictly confidential, will be used only by persons engaged in and for the purposes of the survey, and will not be disclosed or released to others for any other purposes (22 FR 1687).

Form NHS-I (3-18-57)

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS
Acting as Collecting Agent for the
U.S. PUBLIC HEALTH SERVICE

**NATIONAL HEALTH SURVEY**

1. Questionnaire ____ nf ____ Questionnairs

2. (a) Address or description of location

3. Iden. Code | 4. Sub-sample weight | 5. Sample | 6. PSU Number | 7. Segment No. | 8. Serial No.

9. Is this house on a farm or ranch? ☐ Yes ☐ No

(b) Type of living quarters: ☐ Dwelling unit ☐ Other

(c) Name of Special Dwelling Place | Code

10. What is the telephone number here? ☐ No phone

11. What is the best time to call?

12. Are there any other living quarters, occupied nr vacant, in this building (apartment)? ☐ Yes ☐ No

Ask at all units except apartment houses

13. Is there any other building on this property for people to live in - either occupied or vacant? ☐ Yes ☐ No

14. Does anyone else living in this building use YOUR ENTRANCE to get to his living quarters? ☐ Yes ☐ No

**INSTRUCTIONS**

If "Yes" to questions 12, 13 or 14 apply definition of a dwelling unit to determine whether one or more additional questionnaires should be filled and whether the listing is to be corrected.

**15. RECORD OF CALLS AT HOUSEHOLDS**

| Item | | I | Com. | 2 | Com. | 3 | Com. | 4 | Com. | 5 | Com. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Entire household | Date / Time | | | | | | | | | | |
| Callbacks for individual respondents | Col. No. ____ Date / Time | | | | | | | | | | |

**16. REASON FOR NON-INTERVIEW**

| TYPE: | A | B | C | Z |
|---|---|---|---|---|
| Reason: | ☐ Refusal<br>☐ No one at home-repeated calls<br>☐ Temporarily absent<br>☐ Other *(Specify)* | ☐ Vacant - Non-seasonsi<br>☐ Vacant - seasonal<br>☐ Usual residence elsewhere<br>☐ Armed Forces<br>☐ Other *(Specify)* | ☐ Demolished<br>☐ In sample by mistake<br>☐ Eliminated in sub-sample<br>☐ Other *(Specify)* | Interview not obtained for:<br>Cols. ____ ____ ____ ____ ____<br>because: |

Comments on non-interview

17. Signature of Interviewer: 18. Code:

Special instructions or notes

**EDITING RECORD FOR OFFICE USE ONLY**

| a. Result of edit | b. Type of follow-up | d. Edited | e. Re-edited | f. Re-edited |
|---|---|---|---|---|
| ☐ Passed<br>☐ Passed (EFQ)<br>☐ Failed - no follow-up<br>☐ Failed - follow-up | ☐ Office telephone<br>☐ Interviewer telephone<br>☐ Personsi<br>c. Result of follow-up<br>☐ Completed ☐ Non-interview | Editor<br>Date | Editor<br>Date | Editor<br>Date |

I. (a) What is the name of the head nf this household? (Enter name in first column)

(b) What are the names nf all other persons who live here? (List all persons who usually live here, and all persons staying here who have no usual place of residence elsewhere. List these persons in the prescribed order.)

(c) Do any (other) lodgers or roomers live here? ☐ No ☐ Yes (List) →

(d) Is there anyone else who lives here who is now away on business? On a visit? Temporarily in a hospital? ☐ No ☐ Yes (List) →

(e) Is there anyone else staying here now? ☐ No ☐ Yes (List) →

(f) Do any of these people have a home elsewhere? ☐ No (leave on questionnaire) ☐ Yes (If not a household member, delete)

Last name

First name and initial

2. How are you related to the head nf the household? (Enter relationship to head, for example: head, wife, daughter, grandson, mother-in-law, partner, lodger, lodger's wife, etc.)

Relationship

3. Race (Check one box for each person) ☐ White ☐ Negro ☐ Other

4. Sex (Check one box for each person) ☐ Male ☐ Female

5. How old were you on your last birthday? Age ☐ Under 1 year

6. Where were you born? (Record state or foreign country) (State or foreign country)

If 14 years old or over, ask:

7. Are you now married, widowed, divorced, separated nr never married? (Check one box for each person) ☐ Under 14 years ☐ Married ☐ Divorced ☐ Widowed ☐ Separated ☐ Never married

If 14 years old or over, ask:

8. What is the highest grade you completed in school? (Circle highest grade completed or check "None") ☐ None ☐ Under 1st years

Elem: 1 2 3 4 5 6 7 8

High: 1 2 3 4

College: 1 2 3 4 5+

If Male and 14 years old or over, ask:

9. (a) Did you ever serve in the Armed Forces of the United States?
   If "Yes," ask:
   (b) Are you now in the Armed Forces, not counting the reserves?
   (If "Yes," delete this person from questionnaire) →

☐ Pers. of und. 14 yrs.
☐ Yes ☐ No
☐ Yes ☐ No

(c) Was any of your service during a war or was it peace-time only?
If "War," ask:
(d) During which war did you serve?
If "Peace-time" only, ask:

(e) Was any of your service between June 27, 1950 and January 31, 1955?

☐ War ☐ Peace-time only
☐ Spanish American ☐ WW-II
☐ WW - I ☐ Korean
☐ Yes ☐ No

If 6 years old or over, ask:

10. (a) What were you doing most of the past 12 months --
    (For males over 16): working, looking for work, or doing something else?
    (For females over 16): working, looking for work, keeping house, or doing something else?
    (For children 6 - 16): going to school or doing something else?
    If "Something else" checked, and person is 50 years old or over, ask:

    (b) Are you retired?

☐ Under 6 years
☐ Working
☐ Looking for work
☐ Keeping house
☐ Going to school
☐ Something else
☐ Yes ☐ No

**I** Interview each adult person for himself for questions 11-26 and Tables I, II, and A, if he is at home. Enter column number of respondent in each column.

☐ Responded for self
Col. No. _____ was respondent

We are interested in all kinds of illness, whether serious or not --

11. Were you sick at any time LAST WEEK OR THE WEEK BEFORE?
    (a) What was the matter?
    (b) Anything else?

☐ Yes ☐ No

12. Last week or the week before did you have any accidents or injuries, either at home or away from home?
    (a) What were they?
    (b) Anything else?

☐ Yes ☐ No

13. Last week or the week before did you feel any ill effects from an earlier accident or injury?
    (a) What were these effects?
    (b) Anything else?

☐ Yes ☐ No

14. Last week or the week before did you take any medicine or treatment for any condition (besides ... which you told me about)?
    (a) For what conditions?
    (b) Anything else?

☐ Yes ☐ No

15. AT THE PRESENT TIME do you have any ailments or conditions that have continued for a long time? (If "No") Even though they don't bother you all the time?
    (a) What are they?
    (b) Anything else?

☐ Yes ☐ No

16. Has anyone in the family - you, your--, etc. - had any of these conditions DURING THE PAST 12 MONTHS?

    (Read Card A, condition by condition; record any conditions mentioned in the column for the person)

☐ Yes ☐ No

17. Does anyone in the family have any of these conditions? (Read Card B, condition by condition; record any conditions mentioned in the column for the person)

☐ Yes ☐ No

## Table I - ILLNESSES, IMPAIRMENTS AND ACCIDENTS

| Line Number | Col. No. of person (a) | Question No. (b) | Did you ever talk to a doctor about ...? (c) | What did the doctor say it was? -- did he use any medical terms? (If doctor not talked to - "No," in col. (c) - record respondent's description) (If ill-effects of earlier accident also fill Table A) For an accident or injury occurring during past 2 weeks, ask: What part of the body was hurt? What kind of injury was it? Anything else? (Also, fill Table A) (d-1) | If an impairment or symptom, ask: What was the cause of ...? (If cause is already entered in (d-1) circle "X" without asking the question) (If accident or injury, fill Table A) (d-2) | (If eye trouble of any kind and 5 years old or over, ask: Can you read ordinary newspaper print with glasses?) (d-3) | What kind of ...trouble is it? (If kind of trouble already entered in col. (d-1), circle "X" without asking the question) (d-4) | What part of the body was affected? (If part of body can be determined from entries in cols. (d-1) through (d-4), circle "X" without asking the question) (d-5) | LAST WEEK OR THE WEEK BEFORE did ...cause you to cut down on your usual activities for as much as a day? Check one: No / Yes (Go to Col. (f)) (e) (f) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | | ☐ Yes ☐ No | | X | ☐ Yes ☐ No | X | X | |

## Table II - HOSPITALIZATION DURING PAST 12 MONTHS

| Line Number | Col. No. of person (a) | Question No. (b) | When did you enter the hospital? (Month, Year) (c) | How many days were you in the hospital, not counting the day you left? (d) | To interviewer: How many of these--days were in the past 12 months? (e) | To interviewer: How many of these--days were during the past 2 weeks, ending last Sunday? (f) | To interviewer: Was this person still in the hospital last Sunday night? (Verify that no hosp. days after Sunday are in Col. d) (g) | What was the matter? Anything else? (Record each condition in same detail as called for in Table I. If condition is result of accident or injury, also fill Table A) (h) |
|---|---|---|---|---|---|---|---|---|
| 1 | | | No. _____ Year _____ | _____ Days | ☐ All or _____ Days | _____ Days | ☐ Yes ☐ No | |

## TABLE A (Accidents and Injuries)

Line No. from Table I: ☐

1. What part of the body was hurt? What kind of injury was it? Anything else? ☐ Accident happened during past 2 weeks

2. When did it happen? Month _____ Year _____ (Enter only the year if prior to 1956) ☐ Accident happened during past 2 weeks

3. Where did the accident happen?
   ☐ At home (inside or outside the house) ☐ While in Armed Services ☐ Some other place

4. Was a car, truck, bus or other motor vehicle involved in the accident in any way? ☐ Yes ☐ No

5. Were you at work at your job or business when the accident happened? ☐ Yes ☐ No ☐ Under 14 years at time of accident

## MEDICAL CARE

18. (a) LAST WEEK OR THE WEEK BEFORE did anyone in the family - you, your--, etc.- talk to a doctor or go to a doctor's office or clinic? Anyone else?
☐ Yes ☐ No (skip to q. 20)

If "Yes"
(b) How many times during the past 2 weeks? _____ No. of times

(c) Where did you talk to the doctor?
(d) How many times at -- (home, office, clinic, etc.)?
(Record total number of times for each type of place)

| Place | Times |
|---|---|
| At home..... | |
| At office..... | |
| Hospital clinic..... | |
| Company or industry... | |
| Over telephone..... | |
| Other (*Specify*)..... | |

19. What did you have done?
If more than one visit or telephone call:
What did you have done on the {first / second / etc.} visit (or telephone call)?

(1) (2) (3)
☐☐☐ Diag. or treatment
☐☐☐ Pre/post natal care
☐☐☐ Gen'l check-up
☐☐☐ Immun./Vacc.
☐☐☐ Eye exam. (glasses)
☐☐☐ Other (*Specify*)

20. If "No" to q. 18a, ask:
How long has it been since you last talked to a doctor? _____ Wks. or _____ Yrs.
☐ Less than 1 mo. ☐ Never

## DENTAL CARE

21. (a) Last week or the week before did anyone in the family go to a dentist? Anyone else?
☐ Yes ☐ No (skip to q. 22)

If "Yes"
(b) How many times during the past 2 weeks? _____ No. of times

22. What did you have done?
If more than one visit:
What did you have done on the {first / second / etc.} visit?

(1) (2) (3)
☐☐☐ Fillings
☐☐☐ Extractions or other surgery
☐☐☐ Straightening
☐☐☐ Treatment for gums
☐☐☐ Cleaning teeth
☐☐☐ Other (*Specify*)

If "No" to q. 21a, ask:
23. How long has it been since you went to a dentist? _____ Mo. or _____ Yrs.
☐ Less than 1 mo. ☐ Never

24. Is there anyone in the family who has lost all of his teeth? ☐ Yes ☐ No

## HOSPITAL CARE

25. (a) DURING THE PAST 12 MONTHS has anyone in the family been a patient in a hospital overnight or longer? ☐ Yes (Table II) ☐ No

If "Yes":
(b) How many times were you in the hospital? _____ No. of times

26. (a) During the past 12 months has anyone in the family been a patient in a nursing home or sanitarium? ☐ Yes (Table II) ☐ No

If "Yes"
(b) How many times were you in a nursing home or sanitarium? _____ No. of times

27. During the past 12 months in which group did the total income of your family fall, that is, your's, your --'s, etc.? (Show Card H) Include income from all sources, such as wages, salaries, rents from property, pensions, help from relatives, etc. Group No.

## Table I - ILLNESSES, IMPAIRMENTS AND ACCIDENTS

| How many days, including the 2 weekends? | How many of these -- days were you in bed all or most of the day? | If 6 years old or over, ask: Last week or the week before would you have been working at a job or business (going to school) except for ...? | If "Yes" in col. (i): How many days did ... keep you from work (going to school)? | Did you first notice ... DURING THE PAST 3 MONTHS or before that time? Check one: Before 3 months | During 3 months | To Interviewer: If Col. (k) is checked or the condition is on either one of Cards A or B, continue; otherwise, STOP | Did you first notice ... DURING THE PAST 12 MONTHS or before that time? | When did you last talk to a doctor about ...? | Do you still take any medicine or treatment that the doctor prescribed for ...? | About how many days during the past 12 months, has ... kept you in bed for all or most of the day? | Please look at this card and read each statement. Then tell me which statement fits you best. | If "1," or "2" or "3" in Col. (f) ask: Please look at this card and tell me which of these statements fits you best. | Line Number |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (g) | (h) | (i) | (j) | (k) | (l) | (m) | (n) | (o) | (p) | (q) | (r) | (s) | |
| _____ Days | _____ Days or ☐ None | ☐ Yes ☐ No | _____ Days or ☐ None | | | ☐ Last ☐ Before 2 wks. ☐ Week before | Mo. _____ Yr. _____ ☐ Before ☐ Birth | Mo. _____ Yr. _____ ☐ No Dr. | ☐ Yes ☐ No ☐ No Dr. | _____ Days or ☐ None | | | 1 |

## Table II - HOSPITALIZATION DURING PAST 12 MONTHS

| Were any operations performed on you during this stay in the hospital? If "Yes": (a) What was the operation? (b) Any other operations? (1) | What is the name and address of the hospital you were in? (Enter name, city or county, and State) (j) | Line Number |
|---|---|---|
| ☐ Yes → ☐ No | | 1 |

## FOOTNOTES AND COMMENTS

Card A

NATIONAL HEALTH SURVEY

Check List of Chronic Conditions

1. Asthma
2. Any allergy
3. Tuberculosis
4. Chronic bronchitis
5. Repeated attacks of sinus trouble
6. Rheumatic fever
7. Hardening of the arteries
8. High blood pressure
9. Heart trouble
10. Stroke
11. Trouble with varicose veins
12. Hemorrhoids or piles
13. Gallbladder or liver trouble
14. Stomach ulcer
15. Any other chronic stomach trouble
16. Kidney stones or other kidney trouble
17. Arthritis or rheumatism
18. Prostate trouble
19. Diabetes
20. Thyroid trouble or goiter
21. Epilepsy or convulsions of any kind
22. Mental or nervous trouble
23. Repeated trouble with back or spine
24. Tumor or cancer
25 Chronic skin trouble
26. Hernia or rupture

Card B

NATIONAL HEALTH SURVEY

Check List of Impairments

1. Deafness or serious trouble with hearing.
2. Serious trouble with seeing, even with glasses.
3. Condition present since birth, such as cleft palate or club foot.
4. Stammering or other trouble with speech.
5. Missing fingers, hand, or arm.
6. Missing toes, foot, or leg.
7. Cerebral palsy.
8. Paralysis of any kind.
9. Any permanent stiffness or deformity of the foot or leg, fingers, arm, or back.

Card C

NATIONAL HEALTH SURVEY

For:

Workers and other persons except Housewives and Children

1. Cannot work at all at present.
2. Can work but limited in amount or kind of work.
3. Can work but limited in kind or amount of outside activities.
4. Not limited in any of these ways.

Card D

NATIONAL HEALTH SURVEY

For: Housewife

1. Cannot keep house at all at present.
2. Can keep house but limited in amount or kind of housework.
3. Can keep house but limited in outside activities.
4. Not limited in any of these ways.

Card E

NATIONAL HEALTH SURVEY

For:

Children from 6 to 16 years old and others going to school

1. Cannot go to school at all at present time.
2. Can go to school but limited to certain types of schools or in school attendance.
3. Can go to school but limited in other activities.
4. Not limited in any of these ways.

Card F

NATIONAL HEALTH SURVEY

For: Children under 6 years old

1. Cannot take part at all in ordinary play with other children.
2. Can play with other children but limited in amount or kind of play.
4. Not limited in any of these ways.

Card G

NATIONAL HEALTH SURVEY

1. Confined to the house all the time, except in emergencies.
2. Can go outside but need the help of another person in getting around outside.
3. Can go outside alone but have trouble in getting around freely.
4. Not limited in any of these ways.

Card H

NATIONAL HEALTH SURVEY

Family income during past 12 months

1. Under $500 (including loss)
2. $500 - $999
3. $1,000 - $1,999
4. $2,000 - $2,999
5. $3,000 - $3,999
6. $4,000 - $4,999
7. $5,000- $6,999
8. $7,000 - $9,999
9. $10,000 and over.

# APPENDIX II

## ESTIMATING EQUATIONS

In the National Health Household-Interview Survey, the following algebraic statements summarize the estimation process for $X'$ an estimate of X, a population characteristic.

Let $P_{hij}$ be the probability of selecting the $j^{th}$ PSU in the $i^{th}$ stratum in the $h^{th}$ Tab Area, with $P_{hij} = \frac{A_{hij}}{A_{hi}}$, where $A_{hij}$ is 1950 population of the $hij^{th}$ PSU and $A_{hi}$ = 1950 population of the $hi^{th}$ stratum.

and let $A_{chij}$ be number of persons in the $c^{th}$ color-residence group in the $hij^{th}$ PSU according to the 1950 Census.

Then

$$v''_{ch} = \sum_{i=1}^{L_h} \sum_{j=1}^{1} A_{chij} \left(\frac{1}{P_{hij}}\right),$$

where $L_h$ is number of nonself-representing strata in the $h^{th}$ Tab Area, $v''_{ch}$ is an estimate of the number of persons in the $c^{th}$ color-residence group in the nonself-representing strata in the $h^{th}$ Tab Area.

The quantity $v_{ch}$ is the corresponding 1950 Census count. If next, $\dot{X}_{vach}$ is the sample aggregate of the X-measure for the $a^{th}$ age-sex-color class in the $c^{th}$ color-residence group in the nonself-representing strata in the $h^{th}$ Tab Area, and $\dot{X}_{uach}$ is the corresponding aggregate for self-representing strata in the $h^{th}$ Tab Area, and if further $f_h$ is the over-all sampling fraction for the $h^{th}$ Tab Area, then

$$X''_{ah} = \sum_{c} \left[\frac{\dot{X}_{uach}}{f_h} + \frac{\dot{X}_{vach}}{f_h} \frac{v_{ch}}{v''_{ch}}\right]$$

is a first-stage ratio estimate of the characteristic for the $a^{th}$ age-sex-color class in the $h^{th}$ Tab Area.

In precisely the same manner, an estimate of $Z_{ah}$, the current population of the $a^{th}$ age-sex-color class in the $h^{th}$ Tab Area is calculated as

$$\dot{Z}''_{ah} = \sum_{c} \left[\frac{\dot{Z}_{uach}}{f_h} + \frac{\dot{Z}_{vach}}{f_h} \frac{v_{ch}}{v''_{ch}}\right]$$

The total first-stage ratio estimate for the $a^{th}$ age-sex-color class is for the X-measure:

$$X''_a = \sum_{h} X''_{ah}, \text{ and}$$

for population:

$$Z''_a = \sum_{h} Z''_{ah}$$

The final second-stage ratio estimate of the total X-measure is $X' = \sum_{a} \frac{X''_a}{Z''_a} Z_a$, where $Z_a$ is the independent current population estimate for the $a^{th}$ age-sex-color class.

# APPENDIX III

# SAMPLING AND MEASUREMENT ERRORS

## Sampling Error — Basic Formulation

One of the attractive features of probability sampling designs is their inherent quality which permits determination of tolerance limits within which lie findings from the survey. More specifically, for such designs, it may be determined, with any specified degree of confidence, what the maximum differences are between results from the sample and those which would be found in a complete enumeration conducted under identical conditions.

In simple designs, determination of sampling variance most commonly is made in a series of 3 steps: (1) An exact sampling variance formula for the design is derived mathematically in terms of (unknown) population parameters. (2) Sample data for individual units are used to estimate the needed but otherwise unknown parameters. (3) The estimated parameters are substituted into the derived formula, and sampling variances evaluated.

For more complex designs such as that of the health survey, the procedure just outlined is usually not feasible or efficient, even when required formulas have been derived. Different methods, outlined in this section of Appendix III and described in somewhat greater detail in the next, are used in the health survey.

The fundamental rationale of these methods is simple and applies to all probability designs. All observations of a characteristic x are distributed randomly into m groups of k observations each. Each group permits making an estimate of approximately $\frac{1}{m}$ th part of the population total, by a sample design which is essentially the same as the over-all design. Thus if $X'_g$ is the estimate from the $g^{th}$ group, $\overline{X}'$ the mean of the m values $X'_g$, and $X'$ is the over-all estimate, then

$$X' = \sum_{g=1}^{m} X'_g,$$ and the sampling variance of $X'$ is $S^2_{X'} = m\, S^2_{X'_g}$, where $S^2_{X'_g}$ is estimated variance of the group estimates:

$$S^2_{X'_g} = \frac{m}{m-1}\left[\frac{\sum (X'_g)^2}{m} - (\overline{X}')^2\right]$$

This general scheme of estimation has been recognized by a number of statisticians. For example, Deming[4] speaks of it as the Tukey plan; Hansen, Hurwitz, and Madow[2] and others describe it as the random group method. It is being used more widely as electronic computers make it more practicable.

## Sampling Error Functions

The picture just sketched needs to be further highlighted in two important respects. In the health survey, attention usually is centered on estimates of aggregates or on estimates of ratios of two estimated aggregates. In either case, since simple estimated aggregates are obtained as ratios of an estimated statistic to estimated population, the ultimate estimate is a ratio, say $R'$, of two other estimates, say, $Y'$ and $X'$. Under the heading "Estimating Sampling Variances From Survey Data," beginning on page 27, a procedure for determining variance of a quantity $X'$ or $Y'$ is presented. An entirely analogous procedure yields the covariance of $X'$ with $Y'$. Finally, rel-variance of the estimate $R'$ is obtained from the equation:

$$V^2_{R'} = V^2_{X'} + V^2_{Y'} - 2V_{X'Y'},$$

where the V-symbols represent relative variances and covariance of the subscript variables.

Thus the procedure can give variance for any aggregate or ratio. In the health survey, thousands of different estimates are being made. Even with high-speed computers, the cost of calculating variances for each separate estimate would be prohibitive. Further, such a step would be undesirable in that it would yield estimated variances which,

because of their own sampling error, would appear at times to be inconsistent among themselves. For these reasons, either one of two courses is followed in Survey publications. In one of these, variances are calculated for only a few key items, and the reader is allowed to infer from these the order of variance for other items.

In the second course, a group of variables having certain common characteristics—such, for example, as being binomial variates—but differing in absolute size, are used to establish a fitted curve which expresses a "law" of variance for variables of the class. The fitted curve usually takes the form $v^2_{X'} = a + \frac{b}{X'}$, where $X'$ is an estimate, $v^2_{X'}$, its relative sampling variance, and a and b are constants of the fitted curve. It is these readings from the curve which are used as best estimates of the variances.

## Estimating Sampling Variances From Survey Data

For calculation of variances from sample data, the universe is divided into the 4 sectors displayed in table 1. The contribution of each sector to overall variance is computed separately.

For sectors I and III the sampling ratio in the first stage of selection is unity. Accordingly, the between-PSU component of variance for these sectors is zero.

Table 1. Sectors for use in calculating variances for a calendar quarter

| Sector number | Sector name | Number of PSU's | Number of segments |
|---|---|---|---|
| I | Self-representing SMA's------- | 77 | 703 |
| II | Nonself-representing SMA's--- | 13 | 63 |
| III | Self-representing urban and rural PSU's----- | 28 | 90 |
| IV | Nonself representing urban and rural PSU's- | 234 | 692 |

The general scheme of estimating within-PSU variance for these sectors is the random group method previously mentioned. It will be illustrated for sector I, the self-representing Standard Metropolitan Areas.

The segment is the unit of sampling within the PSU's, and accordingly is made the basis of calculations of within-PSU variance. The 703 segments in the sector are divided into 8 groups whose membership of approximately 88 segments each is randomly determined. The selection process is controlled so that each group has its proportionate part of each of the types of segments in the sector. The numbers of segments by types are:

| Type | Number of segments |
|---|---|
| Total---------- | 703 |
| Central City---------- | 354 |
| Urban fringe---------- | 156 |
| Other urban places---------- | 30 |
| Rural---------- | 84 |
| New construction areas------ | 79 |

Inflated totals for a characteristic Y for each group are established, with the summarizing operations:

$$Y'_g = \sum_{i=1}^{k} Y'_{gi}$$

where $Y'_{gi}$ is the estimate for that part of the universe which is represented by the $i^{th}$ segment in the $g^{th}$ group, and k is the number of segments in the $g^{th}$ group. The variances of $Y'_g$ and of $Y'$ are calculated as are those for $X'_g$ and $X'$ respectively on page 26, so that

$$S^2_{Y'} = \frac{1}{m-1}\left[m \sum_{g=1}^{m} (Y'_g)^2 - \left(\sum_{g=1}^{m} Y'_g\right)^2\right]$$

The contribution of sector III is calculated in the same manner.

For the nonself-representing sectors, an ultimate cluster technique is employed in calculating variances. Further, since there is but 1 PSU in each stratum, the strata are grouped into pairs, placing similar strata in the same pair. This process is described as a collapsed strata technique. It is illustrated for the nonself-representing urban and rural PSU's.

Data for the 234 PSU's are consolidated into 117 pairs, $Y'_g$ being the estimated total for that

part of the population represented by the $g^{th}$ pair and $Y'_{gi}$ being that part of the total estimated by the $i^{th}$ PSU in the $g^{th}$ pair. Variance of $Y'_{gi}$ is estimated as

$$S^2_{Y'_{gi}} = \frac{1}{L_g - 1} \sum_{i=1}^{L_g} \left(Y'_{gi} - \frac{P_{gi}}{P_g} Y'_g\right)^2 ,$$

where $P_{gi}$ is 1950 population of $i^{th}$ PSU, $P_g$ is 1950 population of $g^{th}$ group and $L_g$ is the number of PSU's in each group. Since all groups contain 2 PSU's, $L_g$ is a constant equal to 2. Further, since $Y'_g = Y'_{g1} + Y'_{g2}$, the sampling variance of $Y'_g$ is

$$S^2_{Y'_g} = 2 S^2_{Y'_{gi}} ,$$

and the total for the sector, $Y'$, has variance

$$S^2_{Y'} = \sum_{g=1}^{117} S^2_{Y'_g} .$$

The same procedure is used for sector II.

Variance of the survey total is simply the sum of the variances for the 4 sectors.

### Measurement Error

Measurement error can be divided into components in a variety of ways. One useful scheme is to separate it into bias and nonsampling variance. Nonsampling variance has in turn many components. Among these are variations which have their source in respondent, interviewer, classifier, editor, or tabulator. The method of estimating sampling variance which is used in the health survey includes most of the measurement variations, although it does not include those components of variation which are unaffected by the size of the sample. With some exceptions (found in edits for consistency), the biases of measurement, from whatever source, are not treated in the present report.

The main text of the report, on page 19, lists several routes being taken, all intended to improve evaluation of measurement error. An ultimate goal is establishment of a model for analyzing over-all error and its components, and for guidance toward efficient use of resources in minimizing total error.

# APPENDIX IV

## STRATIFICATION OF PRIMARY SAMPLING UNITS

### Principles

A twin objective of many sampling designs is a pattern in which individual primary sampling units are as internally heterogeneous as possible, and in which each stratum formed from grouped PSU's is as homogeneous with respect to PSU's as possible. Said in another way this means that ultimate sampling units within a PSU should tend to be unlike one another, but that PSU's within a stratum should tend to be alike one another. This twin objective was sought in the Health Household-Interview Survey.

Three broad specifications of the survey molded the main outlines of the modes of stratification in the NHS. These were: (1) The requirements of end product which were that separate estimates be prepared for major Standard Metropolitan Areas, for a number of geographic sectors of the country, and for differing densities of population (metropolitan, other urban, and rural). (2) For administrative reasons, and in order to minimize operating costs, stratification in the NHS was to be coincident, insofar as feasible, with that of the Census Bureau's Current Population Survey. (3) The general characterization of the stratifying process was that it produce relatively homogeneous socioeconomic classes of PSU's—with this term being further interpreted to reflect geographic location, density of population, rate of population increase between 1940 and 1950, proportion of nonwhite, type of industry in predominantly urban areas, and type of farming in the rural areas.

Within these specifications, the approximately 1,900 PSU's were classified into 372 strata, the following rules serving as principal further guides in the process. In each case the rule is presented as a positive statement, although obviously there had to be some compromises among rules in order to produce a desirable result.

1. Except where a single PSU was larger than an average stratum—size being measured here as elsewhere in the stratification process by 1950 population—strata were of approximately the same size. This meant about 300,000 persons to a stratum.

2. Since the general design called for sample selection of a single PSU from each stratum with probability proportionate to size, each PSU with the population above a lower cutoff became, by itself, a self-representing stratum. The effect of all rules was to set this cutoff at 400,000 (1950 population).

3. Also included as self-representing or certainty areas were any Standard Metropolitan Areas with the population somewhat less than the cutoff, but within 100 miles of an SMA above the cutoff. The rationale was that the same field organization which served the larger city could also serve the other, and thus reduce costs.

4. Solution of the allocation problem (page 31) led to the conclusion that a nonself-representing Tab Area—that is a Tab Area not made up entirely of self-representing PSU's—should contain not less than 4 sample PSU's if it were a Tab Area of Standard Metropolitan Areas, and not less than 8 sample PSU's otherwise. This meant in turn that such Tab Areas would contain corresponding minimum numbers of strata and this fact influenced ultimately the number of different strata which were formed.

5. Since end-product specifications required, for purposes of comparative analysis, both urban and rural Tab Areas within each geographic section, it was decided to make the first stage of sample selection identical for the other urban and the rural Tab Area within the section. Thus, each PSU drawn from other than Standard Metropolitan Areas became the first-stage unit for 1 urban Tab Area and 1 rural Tab Area, and 2 sets of ultimate stage units or segments—1 for each Tab Area—were drawn from each such PSU. This step had to be taken into consideration later in calculating variances, since first-stage selection for these Tab Areas was not independent.

6. Stratification proceeded in a sequential manner: tentative classification with respect to 1 major specification or rule being followed by tentative subclassification by a second rule and then by further subclassification by a third. As the process continued, occasional changes in the first tentative classifications had to be made. After semifinal stratification was completed, there was a review of results, and a few subjective changes made which reviewers thought would increase socioeconomic homogeneity between PSU's within strata. This introduction of judgment in the stratifying phase of the survey could, of course, produce no bias. If it was well done, it reduced sampling variance; if it was poorly done, at worst it would increase variance.

## Results

As indicated, the principles, specifications, and rules led to a classification of the approximately 1,900 PSU's into 372 strata. Of these, 110 are composed of a single self-representing PSU. Collectively, these 110 strata represent 52 percent of the population in the universe. For them there is no between-PSU component of variance. The remaining 262 strata vary a great deal among one another, some being metropolitan, some urban, some rural, and all obviously exhibiting still other differing features as a consequence of the stratification. Even so, 3 examples of actual strata formed may contribute to a "feel" for the nature of nonself-representing strata in the health survey.

Example A. Sparsely populated stratum

| PSU's (defined by counties) | Preliminary 1950 population |
|---|---|
| Total---------- | 254,235 |
| Coconino, Ariz.---------- | 23,755 |
| Dona Ana, N. Mex.*-------- | 39,044 |
| Graham, Ariz.---------- | 13,018 |
| San Juan, N. Mex.---------- | 18,116 |
| Valencia, N. Mex.---------- | 22,574 |
| Navajo, Ariz.---------- | 29,263 |
| Uintah, Utah---------- | 10,259 |
| Alamosa-Costilla, Colo.--- | 16,572 |
| Mineral-Rio Grande, Colo.---------- | 13,330 |
| Montezuma, Colo.---------- | 9,937 |
| Montrose, Colo.---------- | 15,024 |
| Pinal, Ariz.---------- | 43,343 |

*In each of the three examples, the starred PSU represents the stratum in the sample.

Example B. Moderately densely populated non-Metropolitan stratum

| PSU's (defined by counties) | Preliminary 1950 population |
|---|---|
| Total---------- | 315,623 |
| Harrison-Heard-Troup, Ga.*---------- | 68,008 |
| Florence-Marion, S. C.---- | 112,208 |
| Baldwin-Jones-Twigg, Ga.-- | 45,580 |
| Calendar-Sumter, S. C.---- | 89,827 |

Example C. A nonself-representing SMA stratum

| Standard Metropolitan Area | Preliminary 1950 population |
|---|---|
| Total---------- | 301,706 |
| Springfield, Mo.---------- | 104,118 |
| Sioux City, Iowa---------- | 103,959 |
| St. Joseph, Mo.*---------- | 93,629 |

# APPENDIX V.

# THE SAMPLING ALLOCATION PROBLEM

## Leading Considerations

A fundamental fact which conditions the design of a multipurpose survey and the allocation of resources is that no single factor will determine uniquely the design, but rather a balance must be sought taking into consideration leading objectives. In planning the Health Household-Interview Survey, leading considerations were identified as follows.

1. The survey was expected to provide separate estimates for a number of geographic sections and for metropolitan, urban, and rural sectors. This condition was converted initially to a provision that separate worksheet estimates be produced for each of the defined 41 Tab Areas, although the Tab Areas would be consolidated into a lesser number of groups for most purposes.

2. A household survey was predicated, which in the United States ordinarily means a multistage area design.

3. Tentative determination had been reached as to target sampling tolerances for estimates which were to come from the survey.

4. Preliminary study of requirements and review of probable administrative and operating costs strongly suggested that initially the structure of the health survey should parallel in large measure the Current Population Survey (CPS) which was also a general-purpose survey of households. Significant savings might be possible if the 2 surveys were companion undertakings.

5. The survey was to be a continuing activity, geared to production at quarterly intervals of national estimates of characteristics of high incidence, and production of other statistics for the Nation and for parts of the Nation at annual intervals.

6. Appropriations set budget limitations on the design.

## Outline of Design Solution

The specifications suggested that equal reliability be sought for estimates for each Tab Area. The target tolerances and previous design experience suggested further that a multistage survey could be designed which would meet requirements and which would contain a possible 700 to 1,200 households per year per Tab Area.

Experience with CPS indicated that a total of 300 or more strata with 1 sample PSU in each stratum were desirable. Since the principle had been adopted that the 2 surveys were to be companion activities, and since the CPS was operating with 330 strata, it was decided as a first step to adopt tentatively the CPS stratification for the NHS. This tentative decision was reviewed and modified in a later step.

The budget factor was now introduced. For the tentative design, which was beginning to shape up, it seemed that about 36,000 households per year, or a little under 900 per year per Tab Area, was feasible.

At this point, the precision requirements for each tabulation area were considered in terms of the components of variance. The set of strata for CPS in each tabulation area was examined to see if they were adequate to meet precision requirements for the Tab Areas. In the areas in which the minimum stratum requirements did not appear to be satisfied, additional strata were created, thus bringing more PSU's into the sample. In some cases this was accomplished by splitting an existing stratum into 2 parts, letting the PSU which is in the Current Population Survey represent the part of the stratum in which it falls and selecting a new PSU in the other part. In other cases, it was necessary to rearrange some strata to prevent great variation in strata sizes or in the urban-rural composition of a stratum. In such cases new PSU's were selected, and as a result 69 of the PSU's for the CPS are not included in the NHS. An additional 111 PSU's not in the CPS were selected for the NHS sample.

A principal tool utilized in carrying out the analysis indicated in the previous paragraph is expressed in the approximate relationship

$$V^2_{x'} = \frac{V^2_B}{m} + \frac{V^2_W}{n}, \text{ where}$$

$V^2_B$ is between-PSU rel-variance in the population,

$V^2_W$ is within-PSU rel-variance in the population,

$V^2_{x'}$ is sampling rel-variance of an estimated characteristic

m is the number of PSU's in the sample for a Tab Area, and

n is the number of households in the sample for a Tab Area

Values of $V_B^2$ and $V_W^2$ were calculated for a number of household statistics from the CPS and other surveys. $V_{x'}^2$ and n were set from first appropriations set by joint consideration of target tolerances and budget. For each of the several household statistics a value of m was calculated, using the above equation, for nonself-representing strata. Using "typical" solutions, this step determined the needed number of PSU's in each Tab Area and consequently the number of strata which should be established, since 1 PSU was to be drawn from each stratum.

### Result

The consequence of these actions is the health survey sample design, which was planned to have 372 strata, 372 PSU's, 41 Tab Areas, and 36,000 households with 115,000 persons in it each year.

As noted elsewhere in the report, the original allocation of resources will be modified as consumer interest and experience dictate.

# APPENDIX VI

# ILLUSTRATION OF DRAWING PSU'S AND HOUSEHOLDS INTO THE SAMPLE

## Selection of Primary Sampling Units

Section 5 of this report outlines the main features of sample selection in the health survey. This Appendix illustrates the principal steps of that process.

Assume a particular stratum contains 4 primary sampling units, or PSU's. These are listed, together with their 1950 population, and cumulated population, as in table 2.

Table 2. Primary sampling units in stratum number 428

| PSU | 1950 population[1] | Cumulative 1950 population |
|---|---|---|
| Cedar Rapids, Iowa, SMA---------- | 104,000 | 104,000 |
| Lincoln, Nebr., SMA---- | 118,000 | 222,000 |
| Topeka, Kans., SMA----- | 104,000 | 326,000 |
| Waterloo, Iowa, SMA---- | 99,000 | 425,000 |

[1]Preliminary and approximate population figures are used in this example.

A random number between 1 and 425,000 is selected. Assume the number is 301,265. This number selects Topeka, Kans., as the sample PSU from stratum 428.*

*In three respects, the example is a streamlined version of detailed selection. (1) Where the stratum in the health survey and in the Current Population Survey were identical, the PSU drawn earlier for the CPS was used also in NHS. (2) Where a CPS stratum was divided into 2 strata in NHS, an unbiased selection procedure retained the CPS PSU for one of the new strata. (3) Those PSU's which are found also in CPS were selected initially with probability proportional to size, and also under restrictions of the Goodman-Kish controlled selection technique which increases the probabilities of selection for preferred combinations of units.

## Selection of Enumeration Districts and Segments

The exact procedure for selecting segments varies depending on whether the Tab Area involved is a Standard Metropolitan Area, an "other urban" area, or a rural area, but the nature of the procedure is the same for all areas. It will be described for a typical metropolitan Tab Area for which not all first-stage sampling units were self-representing; i.e., for a Tab Area in which there is more than 1 PSU in the sample. In following this selection process it is useful to remember that the final sample of households and persons is intended to be self-weighting within the Tab Area, which means that every household in the Tab Area has an equal chance of being selected.

Assume that this Tab Area has 5 PSU's in the sample, 3 of which are self-representing, and 2 of which are not. Since the over-all design has an average annual sampling rate of about 1 in 1,400 and since 144 segments are to be selected from the Tab Area, assume this typical Tab Area contains an estimated 200,000 segments in the population (page 13). More precisely, the assumption is that the Tab Area contains 200,000 size measures, where a size measure is equal to 6 households, and the number of size measures is the number of households in 1950 divided by 6.

The first step is to allocate the 144 sample segments to the 5 sample PSU's. This is done in proportion to the estimated size of the stratum represented by the PSU. For example, if a particular sample PSU contains 5,000 size measures, and was drawn from a stratum containing 25,000 size measures, it represents those 25,000 size measures in the sample and, therefore, represents one-eighth part [25,000 divided by 200,000] of the population in the Tab Area. Therefore 1/8 of 144, or 18 segments are assigned to that PSU. In order to facilitate continuous sampling, and to reduce costs by having samples in adjacent quarters also geographically neighboring, 4 quarterly samples are drawn simultaneously, as sketched in the next paragraph. Accordingly, the 18 segments are divided among the 4 quarters, so that either 4 or 5 segments will appear in each quarter.

The next step is to localize the sample into areas smaller than the PSU. For this purpose the enumeration district, or ED, is utilized. ED's, used as administrative and tabulating cells in the 1950 Census, vary greatly in size, but usually contain not less than 10 or more than 150 size measures. Assume in this illustration that the selected PSU with 5,000 size measures contains 50 ED's.

For the PSU of the example, 4 segments will be required in some quarters and 5 in others. The larger of these numbers is identified as the number of "starting points." Thus in this PSU there are 5 starting points. It is intended that these starting points be distributed randomly, but systematically throughout the PSU and that they serve as selectors of ED's and segments for the first quarter. This is done in the following manner. The ED's are arranged in systematic sequence with all central city ED's listed first, followed by all urbanized fringe ED's, and then by other urban ED's, and finally by rural ED's.

The first starting point is determined by choosing a random number between 1 and 1,000 [5,000 size measures in the PSU divided by 5, the number of starting points]. Say this number is 725. Then that listed ED which contains the 725th cumulated size measure is included in the sample, as are also ED's with the 1,725th, 2,725th, 3,725th, and 4,725th cumulated size measures.

Consider the ED with the 725th size measure. Suppose it contained 100 size measures, identified in the cumulated listing as numbers 705 through 804. The process just described locates the first starting point then not only in this particular ED, but at the 21st size measure [random number 725, minus 705, plus 1].

Making use of Sanborn* and other detailed maps, the ED then is "segmented" on a new map into 100 units approximately equal in size (i.e., in the number of expected households). These units are numbered consecutively from 1 through 100 in a systematic fashion beginning with a randomly located start. The unit or segment numbered 21, containing an expected 6 households, becomes a sample segment for the first quarter of interviewing. This same procedure is carried out for other choosen ED's in the PSU and for other sample PSU's in the Tab Area.

It will be noticed that, because some numbers are not exactly divisible by others, in the example PSU 5 rather than the calculated 4.5 segments are interviewed in the first quarter. Memorandum records are maintained so that over the Tab Area exactly 1/4 of 144 or 36 segments are interviewed each quarter.

In the example ED, the 21st segment was interviewed the first quarter. For the second quarter, the $[725 + \frac{1,000}{4}] = 975$th segment, is in the sample; the 1,225th in the third quarter, and the 1,475th in the fourth quarter; except that the memorandum record again is used to assure that only 18 segments from the PSU are included over the year. In the following year, segments are selected in such a manner that they are geographically neighboring the segments in the first sample at about the same time of the year.

Thus it is that over the year, for the stratum from which the example PSU comes, the probability that any segment, household, or person is in the sample is the product of the probability of selecting this particular PSU (5,000 divided by 25,000) times the probability of selecting a particular segment within the PSU (18 divided by 5,000); or in other words is 1/5 times 18/5,000, which is 0.00072. By virtue of the way in which the sample was distributed, this is exactly the designed over-all sampling portion, 144/200,000, for the Tab Area. The probability for any person from the example Tab Area appearing in a given quarter is approximately 0.00018.

## Variations of Detail

The principles of selection were uniform throughout the survey. Depending upon the particular areas which fell into the sample and upon the types of resources available for those areas, additional steps sometimes were taken in the selection process. For example, detailed block statistics were available for many cities. In these cases, a selection of blocks proportional to size was made within sample ED's before making a direct selection of segments. In some instances a block was further subdivided and subsampled before final selection of segments. If it was found from a Sanborn map or other source that the prospective ultimate sampling unit was a large apartment building, still another stage of subsampling was introduced to bring the final unit closer to an expected 6 households.

In some cases, the selection of samples in Washington results in the inclusion of a segment in which the field lister or interviewer finds many more than 6 households. This may occur because of new construction unknown in Washington, or because sampling materials were incomplete or inaccurate. In instances in which the segment obviously appears to contain more than 20 households, field manuals give detailed instructions for subsampling the segment and interviewing only the subsample, in a manner which reduced costs but avoids introduction of bias. A price of slightly higher variance is paid whenever this becomes necessary.

*Published by the Sanborn Map Co., New York, N. Y.

# APPENDIX VII

## RANDOMIZING ASSIGNMENTS, AREAS, AND WEEKS

Basic samples in the health survey are drawn to represent the population of the United States over a calendar quarter. It is efficient in terms of operating procedures and reduction of variance, and furthermore, desirable in terms of potentially available end product, to make each week's collection a random sample of the population. This is done. The randomization of assignments, areas, and weeks is quite an elaborate process. To follow the process through in all its detail most readers would find tedious. For this reason, a description is given by means of an example which exhibits leading features of the process while omitting a number of lesser details.

### Dimensions of the Problem

For administrative reasons (which in the main are consistent with minimum costs) a given interviewer operates within a single geographic section —with a few exceptions—and usually within from 1 to 4 contiguous PSU's. Consequently, randomization of assignments, areas, and weeks was carried out separately within each geographic section. In this process the 8 largest SMA's were excluded from the sections and treated separately.

There are 11 sections in the country, each divided into 3 tab areas: metropolitan, other urban, and rural. Each Tab Area contains 36 segments for the sample for a quarter, and thus a section has 108 segments each quarter. There are a total of 120 interviewers to cover a grand U. S. total of 1,476 segments per quarter (including the 8 largest SMA's). Thus, on the average, 1 interviewer covers 12 segments per quarter. Excluding the 8 largest SMA's, the 108 segments per quarter in a section require an average of 9 interviewers for the section. A typical assignment for an interviewer for a week is 2 segments or an expected 12 households to be interviewed, although an assignment may consist of either 1 or 3 segments. An interviewer may or may not have an assignment in a given week. She never has more than 1 assignment in a week. Thus, the typical situation in a section over a quarter encompasses 54 assignments, 3 tab areas, and 13 weeks, with 6 assignments per interviewer, although the assignments per interviewer may range from 3 to 13. An effort is made to provide at least 1 assignment to each interviewer each month, in order to avoid having too great a time lapse between interviewing experiences.

The objectives of intraquarter arrangements are:

1. Obtaining approximately equal representation from each of the 3 Tab Areas in each section in each week
2. Spacing the work of each interviewer at approximately even intervals over the quarter, and
3. Randomizing assignments (segments to be interviewed) over the weeks of the quarter.

Principal features of the way in which these objectives are reached are illustrated in the following numerical example of a composite geographic section. It should be observed that there is no unique way of accomplishing the objectives and that the method chosen is but one of several possible methods.

### Example

This geographic section contains the usual 3 Tab Areas: metropolitan, urban, and rural, each of which has 36 segments to be interviewed over the quarter. Nine interviewers have been hired for work in the Census Region which contains the section. The Census Regional Offices, of which there are 17, have indicated for each of the interviewers in which of the 20 PSU's in the sample in the section they can serve. This information has been reported to Washington (table 3).

Table 3. Interviewer service areas

| Interviewer | Can serve in PSU(s) numbered |
|---|---|
| A---------- | 1, 2 |
| B---------- | 3 |
| C---------- | 4 |
| D---------- | 5, 6, 7 |
| E---------- | 8, 9 |
| F---------- | 10, 11, 12 |
| G---------- | 13, 14, 15 |
| H---------- | 16, 17, 18 |
| J---------- | 19, 20 |

Step I

Formation of assigments in each PSU.—The sample segments within each PSU are arranged in sequence by degree of urbanization and grouped into assignments of 2 segments each (with 1 assignment containing either 1 or 3 segments if the total number of segments is odd). The purpose of the grouping is to put unlike segments in the same assignments and to obtain balance between urban and rural Tab Areas in the assignments. The process is illustrated in table 4 for a non-SMA primary sampling unit which contains 6 segments; the segments connected by a line being members of the same assignment.

Table 4. Formation of assignments in PSU number 5

| Segment number | Urbanization classification |
|---|---|
| 5-1 | Urban segment |
| 5-6 | Urban segment |
| 5-2 | Rural segment |
| 5-9 | Rural segment |
| 5-14 | Rural segment |
| 5-11 | Rural segment |

Thus 3 assignments are identified for this PSU. In PSU's that are SMA, the arrangement is in sequence by central city-, urban fringe-, other urban places-, and rural-segments.

Step II

Determination of number of assignments for each interviewer.—The number of assignments in each PSU having been determined, the number of assignments for each interviewer is established readily by reference to the field report reflected in table 3. A new worksheet, table 5, is set up combining these two pieces of information. The first figure in each cell is the identification number of the PSU and the second figure is the number of assignments in that PSU. The columns headed total number of assignments, SMA, and non-SMA are utilized later in the allocation process.

Step III

Spacing interviewer assignments throughout the quarter.—The next step is to distribute the number of assignments by week throughout the quarter in such a fashion that each interviewer's work is spaced at approximately even intervals over the quarter and so that the total number of assignments is roughly constant from week to week. This step is carried out on another worksheet shown in table 6.

Table 5. Number of assignments for each interviewer

| Interviewer | Number of assignments by PSU | | | Total number of assignments | | |
|---|---|---|---|---|---|---|
| | | | | All | SMA | Non-SMA |
| A | 1-2 | 2-3 | | 5 | 2 | 3 |
| B | 3-3 | | | 3 | 3 | |
| C[1] | 4-1 | | | 1 | | 1 |
| D | 5-3 | 6-3 | 7-2 | 8 | 3 | 5 |
| E | 8-3 | 9-4 | | 7 | | 7 |
| F | 10-4 | 11-2 | 12-3 | 9 | 3 | 6 |
| G | 13-4 | 14-3 | 15-3 | 10 | 3 | 7 |
| H | 16-2 | 17-3 | 18-2 | 7 | 2 | 5 |
| J | 19-2 | 20-2 | | 4 | 2 | 2 |

[1] Interviewer C has in this example only 1 assignment in the quarter for this section. She has additional assignments in other PSU's in a neighboring section which were assigned because the locations were more accessible to her than to interviewers from the other section.

Table 6. Spacing interviewer assignments by week

| Week number | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| G | G | G | F | G | G | G | F | G | G | G | F | G |
| F | F | D | D | F | F | D | X | F | F | D | X | D |
| D | X | Y | X | D | X | Y | Y | D | X | Y | A | X |
| Y | A | J | A | Y | A | J | A | J | Y | J | B | C |
| | | | B | | | | B | | | | | |

The interviewer with the largest number of assignments—Interviewer G with 10 assignments in this section—has her assignments located by week on the first line of the table. Since she has work in 10 of the 13 weeks, she has an assignment in each week except for 3 evenly spaced and randomly chosen weeks. Note that at this point the identity of each assignment has not been determined, but only the fact that Interviewer G has an assignment in the specified week.

Then the interviewer with the next largest number of assignments—Interviewer F with 9 as-

signments—has her weeks of work posted to table 6. This is done by entering her identification in 9 of the remaining unfilled cells of the table, taking care to fill the first line before starting on the second line and still attempting equal spacing of the 9 assignments. In particular, F is not allowed to have 2 assignments in the same week. This process is continued for each interviewer until the 54 assignments for the section have been placed. The assignment of Interviewer C to week 13 was made with consideration being given also to timing of her assignments in the neighboring section.

Interviewers E and H each have 7 assignments. In table 6 designations X and Y have been used in lieu of E and H without decision as to which is which. This decision is reserved to a later point in order to permit greater flexibility in placing work.

Step IV

Randomizing assignments.—The remaining problem is to match specific assignments randomly with weekly allocations of workload for the interviewers. An important side condition is imposed on this process.

As nearly as possible each week's sample is kept balanced by SMA assignments and non-SMA assignments. In this example, with 54 assignments to be made during the quarter, either 1 or 2 SMA assignments will be made each week, either 2 or 3 non-SMA assignments, and a total of 4 or 5 assignments each week.

Assignments first are made tentatively, and in a few instances it may become necessary for an assignment which has been allocated to one interviewer to be reassigned later in the process to another interviewer as the sequential assignment process reduces degrees of freedom in allocating workloads. Before beginning the randomization of assignments one needs to assemble the data from tables 5 and 6 and from a new table—table 7.

Table 7. Designation of PSU's and assignments as SMA and non-SMA

| Assignments in these PSU's are SMA segments | Assignments in these PSU's are non-SMA segments |
|---|---|
| 1, 3, 6, 12, 15, 16, 19 | 2, 4, 5, 7, 8, 9, 10, 11, 13, 14, 17, 18, 20 |

Assignments within a PSU are then identified by a letter prefixed by a PSU number; e.g., the 3 assignments in PSU Number 5 are 5a, 5b, and 5c.

Table 8 reflects the final allocation and randomization of assignments. The designation in the cell indicates the interviewer and the specific assignment to her in that week. Procedure for filling in the table is outlined in the remaining paragraphs of this Appendix.

The initial determination is number of SMA assignments for each week. As noted earlier, this must be either 1 or 2 for each week. Which weeks get 2 is determined randomly, except that weeks 4 and 8, which are to have a total of 5 assignments, are given 2 SMA assignments each. This action determines also the number of non-SMA assignments for each week and these are posted to table 8.

Allocation is made first then, for the SMA assignment for week 1. Table 6 shows that interviewers G, F, D, and Y are scheduled to work in the first week and the single SMA allotment could be given to any one of the SMA assignments associated with these interviewers. Interviewer Y is not yet identified as to whether she is E or H. Collectively, G, F, D, E, and H account for 11 SMA assignments. One of these is picked at random. The assignment picked was 6b, which also selects interviewer D. The entry D6b is posted in the first cell in week 1.

Two SMA assignments are required for the second week, to be given interviewers G, F, X, or A. The assignments are next selected randomly from the SMA assignments available, as in week 1. The assignments proved to be A1b and F12c. This process is continued for successive weeks.

In drawing for week 6, assignment 16b was selected, and thus X was determined to be H, and Y to be E.

It happened that when week 12 was reached only SMA assignments A1a, B3b, and B3c remained available. Since B could not handle 2 assignments in week 13, A1a was assigned to week 13, along with B3b, which was drawn at random from B3b and B3c. The remaining assignment B3c, went to week 12.

When the SMA assignments had been allocated, the non-SMA allocations were undertaken, beginning with week 1, and using the same procedure as for SMA assignments.

The drawings were such that in the eighth week a non-SMA assignment would have been allotted to interviewer B. However, there was none available to B who had been given all her assignments earlier—she served only SMA territory. Since she had served in lieu of G, D, or H in week 13 for SMA assignment, a random non-SMA assignment from among those still available to G and D was substituted for B in week 8. It turned out to be D7b. Two other similar changes had to be made to complete the panel.

Table 8. Final assignments

| | Week number | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | |
| | D6b | F12c | J19b | F12b | G15c | H16b | D6a | F12a | D6c | F12a | G15a | B3c | A1a | |
| | G13a | A1b | G14b | B3a | E8a | G13d | J19a | H16a | G13b | G14c | J20b | F10a | B3b | |
| | F10c | G13c | D5a | A2a | D7a | F10d | E9b | E8b | F10b | H17a | D5b | H17c | F11b | |
| | E9c | H18a | E8c | H17b | F11a | A2c | G14a | A2b | J20a | E9a | H18b | E9d | C4a | |
| | | | | D5c | | | | D7b | | | | | | All weeks |
| Total SMA Assign | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 18 |
| Total Non-SMA Assign | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 36 |
| Total Assign | 4 | 4 | 4 | 5 | 4 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 4 | 54 |

# APPENDIX VIII

## SELECTED STATISTICS ABOUT THE SURVEY

For ready reference, and for their value in giving quick insight to various features of the health survey, there are assembled in this Appendix several tables of statistics on the survey (tables 9-15). In most instances the figures which are shown are rounded and approximate since they are intended to convey an impression rather than to serve any operational purpose. As a result detailed figures are not always consistent with totals.

Table 9. Summary statistics on components of NHS

| Item | Number |
|---|---|
| Counties and independent cities | 3,100 |
| Primary sampling units in population | 1,900 |
| Primary sampling units in sample | 372 |
| Strata | 372 |
| In national sample in 1 year: | |
| Persons | 115,000 |
| Households | 36,000 |
| Segments | 6,000 |
| Tab Areas | 41 |
| Large SMA's which are separate Tab Areas | 8 |
| Geographic sections | 11 |

Table 10. Size of national sample for different time intervals

| Type of unit | Number of units in | | |
|---|---|---|---|
| | 1 year | 1 quarter | 1 week |
| Persons | 115,000 | 29,000 | 2,200 |
| Households | 36,000 | 9,000 | 700 |
| Segments | 6,000 | 1,500 | 115 |
| PSU's | 372 | 372 | about 60 |

Table 11. Size of sample over 1 year

| Type of unit | Number of units in sample over 1 year | | |
|---|---|---|---|
| | National total | Each geographic section | Each Tab Area |
| Persons | 115,000 | 10,500[1] | 2,800 |
| Households | 36,000 | 3,300[1] | 880 |
| Segments | 6,000 | 550[1] | 145 |
| PSU's | 372 | 34[1] | (2) |

[1] Average.

[2] Urban and rural Tab Areas in a given sample are represented by the same PSU. There is an average of about 18 different sample PSU's for each of the non-SMA, first-stage selections for Tab Areas.

Table 12. Approximate over-all sampling rates on an annual basis

| Sector | Approximate inflation factor (reciprocal of over-all sampling rate) |
|---|---|
| U. S. total | 1,400 |
| New York SMA | 4,700 |
| Chicago SMA | 2,000 |
| Typical other large SMA | 1,000 |
| Tab Area with highest sampling rate | 350 |
| Tab Area with lowest sampling rate (NY) | 4,700 |

Table 13. Data on field supervisors and interviewers

| Item | Amount or number |
|---|---|
| Number of field supervisors | 17 |
| Number of interviewers | 120 |
| Typical interviewer workload in 1 week | 12 households |
| Typical interviewer workload over 1 quarter | 72 households |
| Typical number of interviewers in a geographic section | 9 |
| Typical time required for interview of a household, including travel and call backs (but exclusive of supplemental inquiries) | 60 minutes |

Table 14. Summary operations report on interviewing for 6 months' activity

| Item | Number or percent[1] |
|---|---|
| Number of listings assigned for interview | 24,032 |
| Number of listings demolished, vacant, or otherwise not eligible for interview (Types B and C exclusions) | 3,251 |
| Net number of listings eligible for interview | 20,781 |
| Noninterviews | 1,271 |
| Percent of listings eligible | 6.1 |
| Percent refusal | 1.2 |
| Percent other (not at home, etc.) | 4.9 |
| Number of households with completed interviews | 19,510 |
| Number of persons in households with completed interviews | 62,046 |

[1] Includes approximately 7.5 percent more households than were designed for the basic survey; extra households used in preparing estimates for one part of the country.

Table 15. Primary sampling units by type

| Geographic area | Number of PSU's | | |
|---|---|---|---|
| | Total[1] | Self-representing[1] | Nonself-representing |
| Total | 372 | 110 | 262 |
| Boston SMA | 1 | 1 | - |
| New York SMA | 1 | 1 | - |
| Philadelphia SMA | 1 | 1 | - |
| Pittsburgh SMA | 1 | 1 | - |
| Detroit SMA | 1 | 1 | - |
| Chicago SMA | 1 | 1 | - |
| Los Angeles SMA | 1 | 1 | - |
| San Francisco SMA | 1 | 1 | - |
| Other SMA's | | | |
| Northeast Region | 28 | 21 | 7 |
| North Central Region | 37 | 27 | 10 |
| South Region | 46 | 36 | 10 |
| West Region | 15 | 14 | 1 |
| Other non-SMA PSU's | | | |
| Northeast Region | 34 | 9 | 25 |
| North Central Region | 70 | 1 | 69 |
| South Region | 107 | 1 | 106 |
| West Region | 36 | 2 | 34 |

[1] In detail 9 self-representing PSU's cross section lines and are counted twice.

## REFERENCES


1. U. S. National Health Survey. Origin and Program of the U. S. National Health Survey. Health Statistics. Series A-1. Public Health Service Publication No. 584-A1. Public Health Service, Washington, D. C., May 1958.

2. Hansen, M.H.; Hurwitz, W.N.; and Madow, W.G.: Sample Survey Methods and Theory. Vol. I. John Wiley & Sons, Inc., New York, N. Y., 1953.

3. Cochran, W.G.: Sampling Techniques. John Wiley & Sons, Inc., New York, N. Y., 1953.

4. Deming, W. E.: Some Theory of Sampling. John Wiley & Sons, Inc., New York, N. Y., 1950. p. 96.




☆ U. S. GOVERNMENT PRINTING OFFICE : 1958 O - 474355

## REPORTS FROM THE U. S. NATIONAL HEALTH SURVEY

Series A

1. Origin and Program of the U. S. National Health Survey. Public Health Service Publication No. 584-A1. Price 25 cents.
2. The Statistical Design of the Health Household-Interview Survey. Public Health Service Publication No. 584-A2.

Series B

1. Preliminary Report on Volume of Physician Visits, United States, July-September 1957. Public Health Service Publication No. 584-B1. Price 25 cents.
2. Preliminary Report on Volume of Dental Care, United States, July-September 1957. Public Health Service Publication No. 584-B2. Price 25 cents.
3. Preliminary Report on Number of Persons Injured, United States, July-December 1957. Public Health Service Publication No. 584-B3. Price 30 cents.
4. Preliminary Report on Disability, United States, July-September 1957. Public Health Service Publication No. 584-B4. Price 30 cents.

The Library of Congress Catalog Card

U. S. *National Health Survey.*
The statistical design of the health household-interview survey, by staff of the U. S. National Health Survey and the Bureau of the Census. Washington, U. S. Dept. of Health, Education, and Welfare, Public Health Service, Division of Public Health Methods, 1958.

40 p. illus., map. 26 cm. (*Its* Health statistics, ser. A-2)

(U. S.) Public Health Service. Publication no. 584-A2.
Bibliography: p. 40.

1. Health surveys. 2. U. S.—Statistics, Medical. I. U. S. Bureau of the Census. II. Title: Health household-interview survey. (Series. Series: U. S. Public Health Service. Publication no. 584-A2)

RA11.B15474 no. 2 614.0973 58-61441

Library of Congress

Public Health Service Publication No. 584–A2

For sale by the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C. - Price 35 cents